

P

Access DB# 48171

SEARCH REQUEST FORM

Scientific and Technical Information Center

Requester's Full Name: _____ Examiner #: _____ Date: _____
Art Unit: _____ Phone Number 30 _____ Serial Number: _____
Mail Box and Bldg/Room Location: _____ Results Format Preferred (circle): PAPER DISK E-MAIL

If more than one search is submitted, please prioritize searches in order of need.

Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc, if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention: _____

Inventors (please provide full names): _____

Earliest Priority Filing Date: _____

**For Sequence Searches Only* Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.*

FOR OFFICIAL USE ONLY

STAFF USE ONLY

Searcher: B. Kessner-Lally
Searcher Phone #: 308-4501
Searcher Location: Bited Lab
Date Searcher Picked Up: 8/2/01
Date Completed: 8/6/01
Searcher Prep & Review Time: 5 mi
Clerical Prep Time: 4 mi
Online Time: 2 mi

Type of Search

NA Sequence (#) 2
AA Sequence (#) _____
Structure (#) _____
Bibliographic _____
Litigation _____
Fulltext _____
Patent Family _____
Other _____

Vendors and cost where applicable

STN _____
Dialog _____
Questel/Orbit _____
Dr. Link _____
Lexis/Nexis _____
Sequence Systems CompuLink 1
WWW/Internet _____
Other (specify) _____

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 2, 2001, 16:57:28 : Search time 183.37 Seconds
(without alignments)
178.060 Million cell updates/sec

Title: US-09-701-626A-79

Perfect score: 52

Sequence: 1 tcgcttcgctcactgcggga.....ggcccttaaccaaagtttt 52

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_0601.*

- 1: /cgnl_9/gcgdata/geneseq/geneseq/NA1980.DAT.*
- 2: /cgnl_9/gcgdata/geneseq/geneseq/NA1981.DAT.*
- 3: /cgnl_9/gcgdata/geneseq/geneseq/NA1982.DAT.*
- 4: /cgnl_9/gcgdata/geneseq/geneseq/NA1983.DAT.*
- 5: /cgnl_9/gcgdata/geneseq/geneseq/NA1984.DAT.*
- 6: /cgnl_9/gcgdata/geneseq/geneseq/NA1985.DAT.*
- 7: /cgnl_9/gcgdata/geneseq/geneseq/NA1986.DAT.*
- 8: /cgnl_9/gcgdata/geneseq/geneseq/NA1987.DAT.*
- 9: /cgnl_9/gcgdata/geneseq/geneseq/NA1988.DAT.*
- 10: /cgnl_9/gcgdata/geneseq/geneseq/NA1989.DAT.*
- 11: /cgnl_9/gcgdata/geneseq/geneseq/NA1990.DAT.*
- 12: /cgnl_9/gcgdata/geneseq/geneseq/NA1991.DAT.*
- 13: /cgnl_9/gcgdata/geneseq/geneseq/NA1992.DAT.*
- 14: /cgnl_9/gcgdata/geneseq/geneseq/NA1993.DAT.*
- 15: /cgnl_9/gcgdata/geneseq/geneseq/NA1994.DAT.*
- 16: /cgnl_9/gcgdata/geneseq/geneseq/NA1995.DAT.*
- 17: /cgnl_9/gcgdata/geneseq/geneseq/NA1996.DAT.*
- 18: /cgnl_9/gcgdata/geneseq/geneseq/NA1997.DAT.*
- 19: /cgnl_9/gcgdata/geneseq/geneseq/NA1998.DAT.*
- 20: /cgnl_9/gcgdata/geneseq/geneseq/NA1999.DAT.*
- 21: /cgnl_9/gcgdata/geneseq/geneseq/NA2000.DAT.*
- 22: /cgnl_9/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARTES

Result No.	Score	Query Match	Length	ID	Description
1	52	100.0	52	21	AAZ44968
2	52	100.0	75	21	AAZ88508
3	52	100.0	78	21	AAZ44912
4	52	100.0	78	21	AAZ44927
5	52	100.0	78	21	AAZ44936
6	52	100.0	78	21	AAZ44938
7	52	100.0	78	21	AAZ44940
8	52	100.0	78	21	AAZ44948
9	52	100.0	7300	21	AAZ44982
10	52	100.0	7300	21	AAZ44983
11	52	100.0	13382	21	AAZ44984

12	52	100.0	14143	21	AAZ44981	P. alcaligenes res
13	50.4	96.9	75	21	AAZ88514	P. alcaligenes rep
14	50.4	96.9	78	21	AAZ44956	P. alcaligenes rep
15	49.4	95.0	79	21	AAZ44941	P. alcaligenes rep
16	40	76.9	74	21	AAZ88505	P. alcaligenes rep
17	40	76.9	74	21	AAZ88511	P. alcaligenes rep
18	40	76.9	74	21	AAZ88515	P. alcaligenes rep
19	40	76.9	74	21	AAZ88516	P. alcaligenes rep
20	40	76.9	74	21	AAZ88520	P. alcaligenes rep
21	40	76.9	77	21	AAZ44900	P. alcaligenes rep
22	40	76.9	77	21	AAZ44908	P. alcaligenes rep
23	40	76.9	77	21	AAZ44911	P. alcaligenes rep
24	40	76.9	77	21	AAZ44913	P. alcaligenes rep
25	40	76.9	77	21	AAZ44930	P. alcaligenes rep
26	40	76.9	77	21	AAZ44944	P. alcaligenes rep
27	40	76.9	77	21	AAZ44952	P. alcaligenes rep
28	40	76.9	77	21	AAZ44954	P. alcaligenes rep
29	40	76.9	77	21	AAZ44957	P. alcaligenes rep
30	40	76.9	77	21	AAZ44966	P. alcaligenes rep
31	40	76.9	77	21	AAZ44967	P. alcaligenes rep
32	40	76.9	77	21	AAZ44987	P. alcaligenes rep
33	40	76.9	78	21	AAZ44907	P. alcaligenes rep
34	40	76.9	78	21	AAZ44933	P. alcaligenes rep
35	40	76.9	78	21	AAZ44997	P. alcaligenes rep
36	38.4	73.8	74	21	AAZ88503	P. alcaligenes rep
37	38.4	73.8	77	21	AAZ44894	P. alcaligenes rep
38	38.4	73.8	77	21	AAZ44910	P. alcaligenes rep
39	38.4	73.8	77	21	AAZ44915	P. alcaligenes rep
40	38.4	73.8	77	21	AAZ44929	P. alcaligenes rep
41	38.4	73.8	77	21	AAZ44964	P. alcaligenes rep
42	38.4	73.8	77	21	AAZ44990	P. alcaligenes rep
43	37.4	71.9	77	21	AAZ44961	P. alcaligenes rep
44	37.4	71.9	77	21	AAZ44937	P. alcaligenes rep
45	37	71.2	79	21	AAZ44986	P. alcaligenes rep

ALIGNMENTS

RESULT	1
AAZ44968	ID
AAZ44968 standard; DNA; 52 BP.	AC
AAZ44968;	AC
16-MAY-2000 (first entry)	DT
P. alcaligenes repeat (PAR) element DNA #75.	XX
Diversity-selected gene; restriction enzyme; adhesin; toxin;	DE
detoxifying enzyme; repeat element; PAR; ss.	XX
Pseudomonas alcaligenes.	OS
WO9964632-A1.	XX
16-DEC-1999.	PN
11-JUN-1999; 99WO-US13295.	XX
12-JUN-1998; 98US-0089086.	XX
12-JUN-1998; 98US-0089101.	PR
(NEWE) NEW ENGLAND BIOLABS INC.	XX
Raleigh EA, Vaisvila R, Morgan RD;	PI
WPI; 2000-116558/10.	XX
Cloning intact genes used to isolate genes for restriction enzymes	DR
Claim 10; Page 62; 97pp; English.	XX

CC This invention describes a novel method for cloning intact,
 CC diversity-selected genes (I) from within gene cassettes (GC) which
 CC comprises identifying DNA repeats that flank GC, hybridizing
 CC oligonucleotides (ON) to these repeats and amplification to produce DNA
 CC fragments containing (I), ligating these fragments into a vector and
 CC transforming cells with the vector. This method is used to clone a wide
 CC variety of prokaryotic genes that provide a selective advantage under
 CC particular conditions, particularly those that encode restriction
 CC enzymes (used as reagents in molecular biology); adhesins (for use in
 CC coating or for targeting molecules or organisms to particular sites,
 CC e.g. for competitive exclusion of a selected pathogen); detoxifying
 CC enzymes; toxins that interact with a host, e.g. for synthesis of
 CC inhibitors or antagonists of the toxin, or in vaccination, or a
 CC modification methyltransferase. Intact genes can be cloned directly with
 CC a high probability that the orientation of expression is known in advance
 CC and low probability of association with extraneous, possibly toxic,
 CC genes. AA24894-244980 represent the *Pseudomonas* alcaligenes repeat (PAK)
 CC elements described in the method of the invention.

XX
 SQ Sequence 52 BP; 11 A; 19 C; 12 G; 10 T; 0 other;

Query Match 100.0%; Score 52; DB 21; Length 52;
 Best Local Similarity 100.0%; Pred. No. 2.3e-11;
 Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcgttcgcctcactcggaacggctaaagcgcgcgccttaaccacaaagtta 52
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 1 tcgttcgcctcactcggaacggctaaagcgcgcgccttaaccacaaagtta 52

RESULT 2
 AA288508
 ID AA288508 standard; DNA: 75 BP.
 AC AA288508;
 XX
 DT 16-MAY-2000 (first entry)
 XX
 DE P. alcaligenes repeat (PAR) element DNA PARI5.
 KW Diversity-selected gene; restriction enzyme; adhesin; toxin;
 KW detoxifying enzyme; repeat element; PAK; ss.
 XX
 OS *Pseudomonas* alcaligenes.
 XX
 PN WO9964632-A1.
 XX
 PD 16-DEC-1999.
 XX
 PF 11-JUN-1999; 99WO-US13295.
 XX
 PR 12-JUN-1998; 98US-0089086.
 PR 12-JUN-1998; 98US-0089101.
 XX
 PA (NEW) NEW ENGLAND BIOLABS INC.
 XX
 PI Raleigh EA, Vaisville R, Morgan RD.
 XX
 DR WP1; 2000-116558/10.
 XX
 PT Cloning intact genes used to isolate genes for restriction enzymes -
 XX Example 1B; Fig 3E; 97pp: English.

CC This invention describes a novel method for cloning intact,
 CC diversity-selected genes (I) from within gene cassettes (GC) which
 CC comprises identifying DNA repeats that flank GC, hybridizing
 CC oligonucleotides (ON) to these repeats and amplification to produce DNA
 CC fragments containing (I), ligating these fragments into a vector and
 CC transforming cells with the vector. This method is used to clone a wide
 CC variety of prokaryotic genes that provide a selective advantage under
 CC particular conditions, particularly those that encode restriction

CC enzymes (used as reagents in molecular biology); adhesins (for use in
 CC coating or for targeting molecules or organisms to particular sites,
 CC e.g. for competitive exclusion of a selected pathogen); detoxifying
 CC enzymes; toxins that interact with a host, e.g. for synthesis of
 CC inhibitors or antagonists of the toxin, or in vaccination, or a
 CC modification methyltransferase. Intact genes can be cloned directly with
 CC a high probability that the orientation of expression is known in advance
 CC and low probability of association with extraneous, possibly toxic,
 CC genes. AA288504-288521 represent *Pseudomonas* alcaligenes repeat (PAK)
 CC elements described in the method of the invention.

XX
 SQ Sequence 75 BP; 18 A; 23 C; 17 G; 17 T; 0 other;

Query Match 100.0%; Score 52; DB 21; Length 75;
 Best Local Similarity 100.0%; Pred. No. 2.4e-11;
 Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcgttcgcctcactcggaacggctaaagcgcgcgccttaaccacaaagtta 52
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 22 tcgttcgcctcactcggaacggctaaagcgcgcgccttaaccacaaagtta 73

RESULT 3
 AA244912
 ID AA244912 standard; DNA: 78 BP.
 XX
 AC AA244912;
 XX
 DT 16-MAY-2000 (first entry)
 XX
 DE P. alcaligenes repeat (PAK) element DNA #19.
 XX
 KW Diversity-selected gene; restriction enzyme; adhesin; toxin;
 KW detoxifying enzyme; repeat element; PAK; ss.
 XX
 OS *Pseudomonas* alcaligenes.
 XX
 PN WO9964632-A1.
 XX
 PD 16-DEC-1999.
 XX
 PF 11-JUN-1999; 99WO-US13295.
 XX
 PR 12-JUN-1998; 98US-0089086.
 PR 12-JUN-1998; 98US-0089101.
 XX
 PA (NEW) NEW ENGLAND BIOLABS INC.
 XX
 PI Raleigh EA, Vaisville R, Morgan RD;
 XX
 DR WP1; 2000-116558/10.
 XX
 PT Cloning intact genes used to isolate genes for restriction enzymes -
 XX Claim 7a; Page 59; 97pp: English.

CC This invention describes a novel method for cloning intact,
 CC diversity-selected genes (I) from within gene cassettes (GC) which
 CC comprises identifying DNA repeats that flank GC, hybridizing
 CC oligonucleotides (ON) to these repeats and amplification to produce DNA
 CC fragments containing (I), ligating these fragments into a vector and
 CC transforming cells with the vector. This method is used to clone a wide
 CC variety of prokaryotic genes that provide a selective advantage under
 CC particular conditions, particularly those that encode restriction
 CC enzymes (used as reagents in molecular biology); adhesins (for use in
 CC coating or for targeting molecules or organisms to particular sites,
 CC e.g. for competitive exclusion of a selected pathogen); detoxifying
 CC enzymes; toxins that interact with a host, e.g. for synthesis of
 CC inhibitors or antagonists of the toxin, or in vaccination, or a
 CC modification methyltransferase. Intact genes can be cloned directly with
 CC a high probability that the orientation of expression is known in advance
 CC and low probability of association with extraneous, possibly toxic.

CC genes. AA244894-244980 represent the Pseudomonas alcaligenes repeat (PAR)
 CC elements described in the method of the invention.

XX Sequence 78 BP; 18 A; 24 C; 19 G; 17 T; 0 other;

Query Match 100.0%; Score 52; DB 21; Length 78;
 Best Local Similarity 100.0%; Pred. No. 2.4e-11;
 Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcgcttcgctcactgcggagccgctaaagcgcgccttaaccacacgtta 52
 |||||
 DB 24 tcgcttcgctcactgcggagccgctaaagcgcgccttaaccacacgtta 75

RESULT 4

AA244927
 ID AA244927 standard; DNA; 78 BP.

XX AC AA244927;

XX DT 16-MAY-2000 (first entry)

XX DE P. alcaligenes repeat (PAR) element DNA #34.

XX KW Diversity-selected gene; restriction enzyme; adhesin; toxin;

XX KW detoxifying enzyme; repeat element; PAR; SS.

XX OS Pseudomonas alcaligenes.

XX PN W09964632-A1.

XX PD 16-DEC-1999.

XX PF 11-JUN-1999; 99WO-US13295.

XX PR 12-JUN-1998; 98US-0089086.

XX PR 12-JUN-1998; 98US-0089101.

XX PA (NEWE) NEW ENGLAND BIOLABS INC.

XX PI Raleigh EA, Vaisvila R, Morgan RD;

XX WPI; 2000-116558/10.

XX PT Cloning intact genes used to isolate genes for restriction enzymes

XX PS Claim 7a; Page 59; 97pp; English.

CC This invention describes a novel method for cloning intact,
 CC diversity-selected genes (I) from within gene cassettes (GC) which
 CC comprises identifying DNA repeats that flank GC, hybridizing
 CC oligonucleotides (ON) to these repeats and amplification to produce DNA
 CC fragments containing (I), ligating these fragments into a vector and
 CC transforming cells with the vector. This method is used to clone a wide
 CC variety of prokaryotic genes that provide a selective advantage under
 CC particular conditions, particularly those that encode restriction
 CC enzymes (used as reagents in molecular biology); adhesins (for use in
 CC coating or for targeting molecules or organisms to particular sites,
 CC e.g. for competitive exclusion of a selected pathogen); detoxifying
 CC enzymes; toxins that interact with a host, e.g. for synthesis of
 CC inhibitors or antagonists of the toxin, or in vaccination, or a
 CC modification methyltransferase. Intact genes can be cloned directly with
 CC a high probability that the orientation of expression is known in advance
 CC and low probability of association with extraneous, possibly toxic,
 CC genes. AA244894-244980 represent the Pseudomonas alcaligenes repeat (PAR)
 CC elements described in the method of the invention.

XX Sequence 78 BP; 18 A; 23 C; 18 G; 19 T; 0 other;

Query Match 100.0%; Score 52; DB 21; Length 78;
 Best Local Similarity 100.0%; Pred. No. 2.4e-11;

Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcgcttcgctcactgcggagccgctaaagcgcgccttaaccacacgtta 52
 |||||
 DB 24 tcgcttcgctcactgcggagccgctaaagcgcgccttaaccacacgtta 75

RESULT 5

AA244936
 ID AA244936 standard; DNA; 78 BP.

XX AC AA244936;

XX DT 16-MAY-2000 (first entry)

XX DE P. alcaligenes repeat (PAR) element DNA #43.

XX KW Diversity-selected gene; restriction enzyme; adhesin; toxin;

XX KW detoxifying enzyme; repeat element; PAR; SS.

XX OS Pseudomonas alcaligenes.

XX PN W09964632-A1.

XX PD 16-DEC-1999.

XX PF 11-JUN-1999; 99WO-US13295.

XX PR 12-JUN-1998; 98US-0089086.

XX PR 12-JUN-1998; 98US-0089101.

XX PA (NEWE) NEW ENGLAND BIOLABS INC.

XX PI Raleigh EA, Vaisvila R, Morgan RD;

XX WPI; 2000-116558/10.

XX PT Cloning intact genes used to isolate genes for restriction enzymes

XX PS Claim 7a; Page 60; 97pp; English.

CC This invention describes a novel method for cloning intact,
 CC diversity-selected genes (I) from within gene cassettes (GC) which
 CC comprises identifying DNA repeats that flank GC, hybridizing
 CC oligonucleotides (ON) to these repeats and amplification to produce DNA
 CC fragments containing (I), ligating these fragments into a vector and
 CC transforming cells with the vector. This method is used to clone a wide
 CC variety of prokaryotic genes that provide a selective advantage under
 CC particular conditions, particularly those that encode restriction
 CC enzymes (used as reagents in molecular biology); adhesins (for use in
 CC coating or for targeting molecules or organisms to particular sites,
 CC e.g. for competitive exclusion of a selected pathogen); detoxifying
 CC enzymes; toxins that interact with a host, e.g. for synthesis of
 CC inhibitors or antagonists of the toxin, or in vaccination, or a
 CC modification methyltransferase. Intact genes can be cloned directly with
 CC a high probability that the orientation of expression is known in advance
 CC and low probability of association with extraneous, possibly toxic,
 CC genes. AA244894-244980 represent the Pseudomonas alcaligenes repeat (PAR)
 CC elements described in the method of the invention.

XX Sequence 78 BP; 17 A; 24 C; 19 G; 18 T; 0 other;

Query Match 100.0%; Score 52; DB 21; Length 78;
 Best Local Similarity 100.0%; Pred. No. 2.4e-11;
 Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcgcttcgctcactgcggagccgctaaagcgcgccttaaccacacgtta 52
 |||||
 DB 24 tcgcttcgctcactgcggagccgctaaagcgcgccttaaccacacgtta 75

RESULT 6

AAZ44938
ID AAZ44938 standard; DNA: 78 BP.
AC AAZ44938;
XX
DT 16-MAY-2000 (first entry)
XX
DE P. alcaligenes repeat (PAR) element DNA #45.
XX
KW Diversity-selected gene; restriction enzyme; adhesin; toxin;
KW detoxifying enzyme; repeat element; PAR; ss.
XX
OS Pseudomonas alcaligenes.
PN WO9964632-A1.
XX
PD 16-DEC-1999.
XX
PF 11-JUN-1999; 99WO-US13295.
XX
PR 12-JUN-1998; 98US-0089086.
PR 12-JUN-1998; 98US-0089101.
XX
PA (NEWF) NEW ENGLAND BIOLABS INC.
XX
PI Raleigh EA, Vaisvila R, Morgan RD;
DR WPI: 2000-116558/10.
XX
PT Cloning intact genes used to isolate genes for restriction enzymes
PS Claim 7a; Page 60; 97pp; English.
XX
CC This invention describes a novel method for cloning intact,
CC diversity-selected genes (I) from within gene cassettes (GC) which
CC comprises identifying DNA repeats that flank GC, hybridizing
CC oligonucleotides (ON) to these repeats and amplification to produce DNA
CC fragments containing (I), ligating these fragments into a vector and
CC transforming cells with the vector. This method is used to clone a wide
CC variety of prokaryotic genes that provide a selective advantage under
CC particular conditions, particularly those that encode restriction
CC enzymes (used as reagents in molecular biology); adhesins (for use in
CC coating or for targeting molecules or organisms to particular sites,
CC e.g. for competitive exclusion of a selected pathogen); detoxifying
CC enzymes; toxins that interact with a host, e.g. for synthesis of
CC inhibitors or antagonists of the toxin, or in vaccination, or a
CC modification methyltransferase. Intact genes can be cloned directly with
CC a high probability that the orientation of expression is known in advance
CC and low probability of association with extraneous, possibly toxic,
CC genes. AAZ44894-244980 represent the Pseudomonas alcaligenes repeat (PAR)
CC elements described in the method of the invention.
XX
SQ Sequence 78 BP; 17 A; 26 C; 18 G; 17 T; 0 other;

Query Match 100.0%; Score 52; DB 21; Length 78;
Best Local Similarity 100.0%; Pred. No. 2.4e-11;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tcgttcgtcactcgggaccggtctaaagcggcccttaacccaacgtta 52
Db 24 tcgttcgtcactcgggaccggtctaaagcggcccttaacccaacgtta 75

RESULT 7
AAZ44940
ID AAZ44940 standard; DNA: 78 BP.
XX
AC AAZ44940;
XX
DT 16-MAY-2000 (first entry)
XX
DE P. alcaligenes repeat (PAR) element DNA #47.

XX
KW Diversity-selected gene; restriction enzyme; adhesin; toxin;
KW detoxifying enzyme; repeat element; PAR; ss.
XX
OS Pseudomonas alcaligenes.
PN WO9964632-A1.
XX
PD 16-DEC-1999.
XX
PF 11-JUN-1999; 99WO-US13295.
XX
PR 12-JUN-1998; 98US-0089086.
PR 12-JUN-1998; 98US-0089101.
XX
PA (NEWF) NEW ENGLAND BIOLABS INC.
XX
PI Raleigh EA, Vaisvila R, Morgan RD;
DR WPI: 2000-116558/10.
XX
PT Cloning intact genes used to isolate genes for restriction enzymes
PS Claim 7a; Page 60; 97pp; English.
XX
CC This invention describes a novel method for cloning intact,
CC diversity-selected genes (I) from within gene cassettes (GC) which
CC comprises identifying DNA repeats that flank GC, hybridizing
CC oligonucleotides (ON) to these repeats and amplification to produce DNA
CC fragments containing (I), ligating these fragments into a vector and
CC transforming cells with the vector. This method is used to clone a wide
CC variety of prokaryotic genes that provide a selective advantage under
CC particular conditions, particularly those that encode restriction
CC enzymes (used as reagents in molecular biology); adhesins (for use in
CC coating or for targeting molecules or organisms to particular sites,
CC e.g. for competitive exclusion of a selected pathogen); detoxifying
CC enzymes; toxins that interact with a host, e.g. for synthesis of
CC inhibitors or antagonists of the toxin, or in vaccination, or a
CC modification methyltransferase. Intact genes can be cloned directly with
CC a high probability that the orientation of expression is known in advance
CC and low probability of association with extraneous, possibly toxic,
CC genes. AAZ44894-244980 represent the Pseudomonas alcaligenes repeat (PAR)
CC elements described in the method of the invention.
XX
SQ Sequence 78 BP; 17 A; 25 C; 20 G; 16 T; 0 other;

Query Match 100.0%; Score 52; DB 21; Length 78;
Best Local Similarity 100.0%; Pred. No. 2.4e-11;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tcgttcgtcactcgggaccggtctaaagcggcccttaacccaacgtta 52
Db 24 tcgttcgtcactcgggaccggtctaaagcggcccttaacccaacgtta 75

RESULT 8
AAZ44948
ID AAZ44948 standard; DNA: 78 BP.
XX
AC AAZ44948;
XX
DT 16-MAY-2000 (first entry)
XX
DE P. alcaligenes repeat (PAR) element DNA #55.
XX
KW Diversity-selected gene; restriction enzyme; adhesin; toxin;
KW detoxifying enzyme; repeat element; PAR; ss.
XX
OS Pseudomonas alcaligenes.
PN WO9964632-A

16-DEC-1999.
11-JUN-1999; 99WO-US13295.
12-JUN-1998; 98US-0089086.
12-JUN-1998; 98US-0089101.
(NEWE) NEW ENGLAND BIOLABS INC.
Kaleigh EA, Vaisvila R, Morgan RD;
WPI: 2000-116558/10.
Cloning intact genes used to isolate genes for restriction enzymes -
Claim 7a; Page 60; 97pp; English.
This invention describes a novel method for cloning intact,
diversity-selected genes (I) from within gene cassettes (GC) which
comprises identifying DNA repeats that flank GC, hybridizing
oligonucleotides (ON) to these repeats and amplification to produce DNA
fragments containing (I), ligating these fragments into a vector and
transforming cells with the vector. This method is used to clone a wide
variety of prokaryotic genes that provide a selective advantage under
particular conditions, particularly those that encode restriction
enzymes (used as reagents in molecular biology); adhesins (for use in
coating or for targeting molecules or organisms to particular sites,
e.g. for competitive exclusion of a selected pathogen); detoxifying
enzymes; toxins that interact with a host, e.g. for vaccination, or a
modification methyltransferase. Intact genes can be cloned directly with
a high probability that the orientation of expression is known in advance
and low probability of association with extraneous, possibly toxic,
genes. AA244894-244980 represent the Pseudomonas alcaligenes repeat (PAR)
elements described in the method of the invention.
XX
SQ Sequence 78 BP; 19 A; 24 C; 18 G; 17 T; 0 other;
Query Match 100.0%; Score 52; DB 21; Length 78;
Best Local Similarity 100.0%; Pred. No. 2.4e-11;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 tcgcttcgctcactcgaggaccggtctaaagccggcccttaacaaacgtta 52
|||||
Db 24 tcgcttcgctcactcgaggaccggtctaaagccggcccttaacaaacgtta 75
RESULT 9
AA244982
ID AA244982 standard; DNA; 7300 BP.
XX
AC AA244982;
XX
DT 16-MAY-2000 (first entry)
XX
DE P. alcaligenes restriction enzyme PacI superintegron DNA #2.
XX
KW Diversity-selected gene; restriction enzyme; adhesin; toxin;
KW detoxifying enzyme; PacI; superintegron; ss.
XX
OS Pseudomonas alcaligenes.
XX
PN WO9964632-A1.
XX
PD 16-DEC-1999.
XX
PF 11-JUN-1999; 99WO-US13295.
XX
PR 12-JUN-1998; 98US-0089086.
PR 12-JUN-1998; 98US-0089101.
XX
PA (NEWE) NEW ENGLAND BIOLABS INC.

XX
PI Kaleigh EA, Vaisvila R, Morgan RD;
XX
DR WPI: 2000-116558/10.
XX
PT Cloning intact genes used to isolate genes for restriction enzymes -
XX
PS Example 1A; Fig 3B-1; 97pp; English.
XX
CC This invention describes a novel method for cloning intact,
CC diversity-selected genes (I) from within gene cassettes (GC) which
CC comprises identifying DNA repeats that flank GC, hybridizing
CC oligonucleotides (ON) to these repeats and amplification to produce DNA
CC fragments containing (I), ligating these fragments into a vector and
CC transforming cells with the vector. This method is used to clone a wide
CC variety of prokaryotic genes that provide a selective advantage under
CC particular conditions, particularly those that encode restriction
CC enzymes (used as reagents in molecular biology); adhesins (for use in
CC coating or for targeting molecules or organisms to particular sites,
CC for competitive exclusion of a selected pathogen); detoxifying
CC toxins that interact with a host, e.g. for vaccination, or a
CC antagonists of the toxin, or in vaccination, or a modification
CC methyltransferase. Intact genes can be cloned directly with a high
CC probability that the orientation of expression is known in advance and
CC low probability of association with extraneous, possibly toxic, genes.
CC This sequence represents a Pseudomonas alcaligenes PacI superintegron
CC DNA sequence.
XX
SQ Sequence 7300 BP; 2023 A; 1944 C; 1641 G; 1690 T; 2 other;
Query Match 100.0%; Score 52; DB 21; Length 7300;
Best Local Similarity 100.0%; Pred. No. 4.7e-11;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 tcgcttcgctcactcgaggaccggtctaaagccggcccttaacaaacgtta 52
|||||
Db 3300 tcgcttcgctcactcgaggaccggtctaaagccggcccttaacaaacgtta 3351
RESULT 10
AA244983
ID AA244983 standard; DNA; 7300 BP.
XX
AC AA244983;
XX
DT 16-MAY-2000 (first entry)
XX
DE P. alcaligenes restriction enzyme PacI superintegron DNA #3.
XX
KW Diversity-selected gene; restriction enzyme; adhesin; toxin;
KW detoxifying enzyme; PacI; superintegron; ss.
XX
OS Pseudomonas alcaligenes.
XX
PN WO9964632-A1.
XX
PD 16-DEC-1999.
XX
PF 11-JUN-1999; 99WO-US13295.
XX
PR 12-JUN-1998; 98US-0089086.
PR 12-JUN-1998; 98US-0089101.
XX
PA (NEWE) NEW ENGLAND BIOLABS INC.
XX
PI Kaleigh EA, Vaisvila R, Morgan RD;
XX
DR WPI: 2000-116558/10.
XX
PT Cloning intact genes used to isolate genes for restriction enzymes -
XX
PS Example 1A; Fig 3C-1; 97pp; English.

XX This invention describes a novel method for cloning intact,
 CC diversity-selected genes (I) from within gene cassettes (GC) which
 CC comprises identifying DNA repeats that flank GC, hybridizing
 CC oligonucleotides (ON) to these repeats and amplification to produce DNA
 CC fragments containing (I), ligating these fragments into a vector and
 CC transforming cells with the vector. This method is used to clone a wide
 CC variety of prokaryotic genes that provide a selective advantage under
 CC particular conditions, particularly those that encode restriction
 CC enzymes (used as reagents in molecular biology); adhesins (for use in
 CC coating or for targeting molecules or organisms to particular sites, e.g.
 CC for competitive exclusion of a selected pathogen); detoxifying enzymes;
 CC toxins that interact with a host, e.g. for synthesis of inhibitors or
 CC methyltransferase. Intact genes can be cloned directly with a high
 CC probability that the orientation of expression is known in advance and
 CC low probability of association with extraneous, possibly toxic, genes.
 CC This sequence represents a Pseudomonas alcaligenes PacI superintegron
 CC DNA sequence.

XX
 SQ Sequence 7300 BP; 2023 A; 1944 C; 1641 G; 1690 T; 2 other;

Query Match 100.0%; Score 52; DB 21; Length 7300;
 Best Local Similarity 100.0%; Pred. No. 4.7e-11;
 Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcgcttcgctcactgcggaccgctaaagccgccccttaacccaaacgtta 52
 |||||
 DB 3300 tcgcttcgctcactgcggaccgctaaagccgccccttaacccaaacgtta 3351

RESULT 11
 AA244984
 ID AA244984 standard; DNA; 13382 BP.
 XX
 AC AA244984;
 XX
 DT 16-MAY-2000 (first entry)
 XX
 DE P. alcaligenes restriction enzyme PacI superintegron DNA #4.
 XX
 KW Diversity-selected gene; restriction enzyme; adhesin; toxin;
 KW detoxifying enzyme; PacI; superintegron; ss.
 XX
 OS Pseudomonas alcaligenes.
 XX
 PN WO9964632-A1.
 XX
 PD 16-DEC-1999.
 XX
 PF 11-JUN-1999; 99WO-US13295.
 XX
 PR 12-JUN-1998; 98US-0089086.
 PR 12-JUN-1998; 98US-0089101.
 XX
 PA (NEW) NEW ENGLAND BIOLABS INC.
 XX
 PI Raleigh EA, Vaisvila R, Morgan RD;
 XX
 DR WPI; 2000-116558/10.
 XX
 PT Cloning intact genes used to isolate genes for restriction enzymes
 XX
 PS Example 1A; Fig 3D-1; 97pp; English.
 XX
 CC This invention describes a novel method for cloning intact,
 CC diversity-selected genes (I) from within gene cassettes (GC) which
 CC comprises identifying DNA repeats that flank GC, hybridizing
 CC oligonucleotides (ON) to these repeats and amplification to produce DNA
 CC fragments containing (I), ligating these fragments into a vector and
 CC transforming cells with the vector. This method is used to clone a wide
 CC variety of prokaryotic genes that provide a selective advantage under

CC particular conditions, particularly those that encode restriction
 CC enzymes (used as reagents in molecular biology); adhesins (for use in
 CC coating or for targeting molecules or organisms to particular sites, e.g.
 CC for competitive exclusion of a selected pathogen); detoxifying enzymes;
 CC toxins that interact with a host, e.g. for synthesis of inhibitors or
 CC antagonists of the toxin, or in vaccination, or a modification
 CC methyltransferase. Intact genes can be cloned directly with a high
 CC probability that the orientation of expression is known in advance and
 CC low probability of association with extraneous, possibly toxic, genes.
 CC This sequence represents a Pseudomonas alcaligenes PacI superintegron
 CC DNA sequence.

XX
 SQ Sequence 13382 BP; 3325 A; 3756 C; 3464 G; 2837 T; 0 other;

Query Match 100.0%; Score 52; DB 21; Length 13382;
 Best Local Similarity 100.0%; Pred. No. 5.2e-11;
 Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcgcttcgctcactgcggaccgctaaagccgccccttaacccaaacgtta 52
 |||||
 DB 8270 tcgcttcgctcactgcggaccgctaaagccgccccttaacccaaacgtta 8321

RESULT 12
 AA244981
 ID AA244981 standard; DNA; 14143 BP.
 XX
 AC AA244981;
 XX
 DT 16-MAY-2000 (first entry)
 XX
 DE P. alcaligenes restriction enzyme PacI superintegron DNA #1.
 XX
 KW Diversity-selected gene; restriction enzyme; adhesin; toxin;
 KW detoxifying enzyme; PacI; superintegron; ss.
 XX
 OS Pseudomonas alcaligenes.
 XX
 PN WO9964632-A1.
 XX
 PD 16-DEC-1999.
 XX
 PF 11-JUN-1999; 99WO-US13295.
 XX
 PR 12-JUN-1998; 98US-0089086.
 PR 12-JUN-1998; 98US-0089101.
 XX
 PA (NEW) NEW ENGLAND BIOLABS INC.
 XX
 PI Raleigh EA, Vaisvila R, Morgan RD;
 XX
 DR WPI; 2000-116558/10.
 XX
 PT Cloning intact genes used to isolate genes for restriction enzymes
 XX
 PS Example 1A; Fig 3A-1; 97pp; English.
 XX
 CC This invention describes a novel method for cloning intact,
 CC diversity-selected genes (I) from within gene cassettes (GC) which
 CC comprises identifying DNA repeats that flank GC, hybridizing
 CC oligonucleotides (ON) to these repeats and amplification to produce DNA
 CC fragments containing (I), ligating these fragments into a vector and
 CC transforming cells with the vector. This method is used to clone a wide
 CC variety of prokaryotic genes that provide a selective advantage under
 CC particular conditions, particularly those that encode restriction
 CC enzymes (used as reagents in molecular biology); adhesins (for use in
 CC coating or for targeting molecules or organisms to particular sites, e.g.
 CC for competitive exclusion of a selected pathogen); detoxifying enzymes;
 CC toxins that interact with a host, e.g. for synthesis of inhibitors or
 CC antagonists of the toxin, or in vaccination, or a modification
 CC methyltransferase. Intact genes can be cloned directly with a high
 CC probability that the orientation of expression is known in advance and

CC low probability of association with extraneous, possibly toxic, genes.
 CC This sequence represents a Pseudomonas alcaligenes PacI superintegron
 CC DNA sequence.

XX SQ Sequence 14143 BP; 3936 A; 3446 C; 3311 G; 3449 T; 1 other;

Query Match 100.0%; Score 52; DB 21; Length 14143;
 Best Local Similarity 100.0%; Pred. No. 5.2e-11;
 Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 tcgttcgctcactgcgggaccggtctaaagcggcccttaaccacagttta 52
 |||||
 Db 4112 tcgttcgctcactgcgggaccggtctaaagcggcccttaaccacagttta 4163

RESULT 13

AAZ88514
 ID AAZ88514 standard; DNA; 75 BP.

XX AAZ88514;

AC AAZ88514;

XX 16-MAY-2000 (first entry)

XX P. alcaligenes repeat (PAR) element DNA PARf11.

XX Diversity-selected gene; restriction enzyme; adhesin; toxin;

KW detoxifying enzyme; repeat element; PAR; ss.

XX Pseudomonas alcaligenes.

OS WO9964632-A1.

PN 16-DEC-1999.

XX 11-JUN-1999; 99WO-US13295.

XX 12-JUN-1998; 98US-0089086.

XX 12-JUN-1998; 98US-0089101.

XX (NEW) NEW ENGLAND BIOLABS INC.

XX Raleigh EA, Vaisvila R, Morgan RD;

XX WPI; 2000-116558/10.

XX Cloning intact genes used to isolate genes for restriction enzymes -

XX Example 1B; Fig 3E; 97pp; English.

CC This invention describes a novel method for cloning intact,
 CC diversity-selected genes (I) from within gene cassettes (GC) which
 CC comprises identifying DNA repeats that flank GC, hybridizing
 CC oligonucleotides (ON) to these repeats and amplification to produce DNA
 CC fragments containing (I), ligating these fragments into a vector and
 CC transforming cells with the vector. This method is used to clone a wide
 CC variety of prokaryotic genes that provide a selective advantage under
 CC particular conditions, particularly those that encode restriction
 CC enzymes (used as reagents in molecular biology); adhesins (for use in
 CC coating or for targeting molecules or organisms to particular sites,
 CC e.g. for competitive exclusion of a selected pathogen); detoxifying
 CC enzymes; toxins that interact with a host, e.g. for synthesis of
 CC inhibitors or antagonists of the toxin, or in vaccination, or a
 CC modification methyltransferase. Intact genes can be cloned directly with
 CC a high probability of association with extraneous, possibly toxic,
 CC genes. AAZ88504-288521 represent Pseudomonas alcaligenes repeat (PAR)
 CC elements described in the method of the invention.

XX SQ Sequence 75 BP; 16 A; 23 C; 19 G; 17 T; 0 other;

Query Match

96.9%; Score 50.4; DB 21; Length 75;

Best Local Similarity 98.1%; Pred. No. 9.9e-11;
 Matches 51; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 tcgttcgctcactgcgggaccggtctaaagcggcccttaaccacagttta 52
 |||||
 Db 22 tcgttcgctcactgcgggaccggtctaaagcggcccttaaccacagttta 73

RESULT 14

AAZ44956

ID AAZ44956 standard; DNA; 78 BP.

XX AAZ44956;

AC AAZ44956;

XX 16-MAY-2000 (first entry)

XX P. alcaligenes repeat (PAR) element DNA #63.

XX Diversity-selected gene; restriction enzyme; adhesin; toxin;

KW detoxifying enzyme; repeat element; PAR; ss.

XX Pseudomonas alcaligenes.

OS WO9964632-A1.

PN 16-DEC-1999.

XX 11-JUN-1999; 99WO-US13295.

XX 12-JUN-1998; 98US-0089086.

XX 12-JUN-1998; 98US-0089101.

XX (NEW) NEW ENGLAND BIOLABS INC.

XX Raleigh EA, Vaisvila R, Morgan RD;

XX WPI; 2000-116558/10.

XX Cloning intact genes used to isolate genes for restriction enzymes -

XX Claim 7a; Page 60; 97pp; English.

CC This invention describes a novel method for cloning intact,
 CC diversity-selected genes (I) from within gene cassettes (GC) which
 CC comprises identifying DNA repeats that flank GC, hybridizing
 CC oligonucleotides (ON) to these repeats and amplification to produce DNA
 CC fragments containing (I), ligating these fragments into a vector and
 CC transforming cells with the vector. This method is used to clone a wide
 CC variety of prokaryotic genes that provide a selective advantage under
 CC particular conditions, particularly those that encode restriction
 CC enzymes (used as reagents in molecular biology); adhesins (for use in
 CC coating or for targeting molecules or organisms to particular sites,
 CC e.g. for competitive exclusion of a selected pathogen); detoxifying
 CC enzymes; toxins that interact with a host, e.g. for synthesis of
 CC inhibitors or antagonists of the toxin, or in vaccination, or a
 CC modification methyltransferase. Intact genes can be cloned directly with
 CC a high probability of association with extraneous, possibly toxic,
 CC genes. AAZ44894-244980 represent the Pseudomonas alcaligenes repeat (PAR)
 CC elements described in the method of the invention.

XX SQ Sequence 78 BP; 16 A; 25 C; 19 G; 18 T; 0 other;

Query Match

96.9%; Score 50.4; DB 21; Length 78;

Best Local Similarity 98.1%; Pred. No. 1e-10;

Matches 51; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 tcgttcgctcactgcgggaccggtctaaagcggcccttaaccacagttta 52
 |||||
 Db 24 tcgttcgctcactgcgggaccggtctaaagcggcccttaaccacagttta 75

```

RESULT 15
AAZ44941
ID AAZ44941 standard; DNA; 79 BP.
XX
XX AC AAZ44941;
XX
XX DT 16-MAY-2000 (first entry)
XX
XX DE P. alcaligenes repeat (PAR) element DNA #48.
XX
XX KW Diversity-selected gene; restriction enzyme; adhesin; toxin;
XX
XX RW detoxifying enzyme; repeat element; PAR; ss.
XX
XX OS Pseudomonas alcaligenes.
XX
XX PN WO9964632-A1.
XX
XX PD 16-DEC-1999.
XX
XX PF 11-JUN-1999; 99WU-US13295.
XX
XX PR 12-JUN-1998; 98US-0089086.
XX
XX PR 12-JUN-1998; 98US-0089101.
XX
XX PA (NFWF ) NEW ENGLAND BIOLABS INC.
XX
XX PI Raleigh EA, Vaisvila R, Morgan RD;
XX
XX DR WPI: 2000-116558/10.
XX
XX PT Cloning intact genes used to isolate genes for restriction enzymes -
XX
XX PS Claim 7a; Page 60; 97pp; English.
XX
XX CC This invention describes a novel method for cloning intact,
XX
XX CC diversity-selected genes (I) from within gene cassettes (GC) which
XX
XX CC comprises identifying DNA repeats that flank GC, hybridizing
XX
XX CC oligonucleotides (ON) to these repeats and amplification to produce DNA
XX
XX CC fragments containing (I), ligating these fragments into a vector and
XX
XX CC transforming cells with the vector. This method is used to clone a wide
XX
XX CC variety of prokaryotic genes that provide a selective advantage under
XX
XX CC particular conditions, particularly those that encode restriction
XX
XX CC enzymes (used as reagents in molecular biology); adhesins (for use in
XX
XX CC coating or for targeting molecules or organisms to particular sites,
XX
XX CC e.g. for competitive exclusion of a selected pathogen); detoxifying
XX
XX CC enzymes; toxins that interact with a host, e.g. for synthesis of
XX
XX CC inhibitors or antagonists of the toxin, or in vaccination, or a
XX
XX CC modification methyltransferase. Intact genes can be cloned directly with
XX
XX CC a high probability that the orientation of expression is known in advance
XX
XX CC and low probability of association with extraneous, possibly toxic,
XX
XX CC genes. AAZ44894-244980 represent the Pseudomonas alcaligenes repeat (PAR)
XX
XX CC elements described in the method of the invention.
XX
XX SQ Sequence 79 BP; 17 A; 27 C; 18 G; 17 T; 0 other;

Query Match 95.0%; Score 49.4; DB 21; Length 79;
Best Local Similarity 98.0%; Pred. No. 2.4e10;
Matches 50; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 cgcttcgctcactcgggacgcggctaaagcgcgcgccttaaccacagtta 52
|||||
Db 26 cgcttcgctcactcgggacgcggctaatgcgcgccttaaccacagtta 76
|||||

Search completed: August 2, 2001, 16:57:31
Job time: 261 sec

```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 2, 2001, 18:40:35 ; Search time 89.09 Seconds
(without alignments)
110.497 Million cell updates/sec

Title: US-09-701-626A-79
Perfect score: 52
Sequence: 1 tcgcttcgctcactcggga.....ggcccttaaccacaaacytta 52

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 324599 seqs, 94655562 residues

Total number of hits satisfying chosen parameters: 649198

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

```
Database : Issued Patents, NA.*
1: /cqn1_7/ptodata/1/ina/5A-COMB.seq.*
2: /cqn1_7/ptodata/1/ina/5B-COMB.seq.*
3: /cqn1_7/ptodata/1/ina/6A-COMB.seq.*
4: /cqn1_7/ptodata/1/ina/6B-COMB.seq.*
5: /cqn1_7/ptodata/1/ina/PCRUS-COMB.seq.*
6: /cqn1_7/ptodata/1/ina/backfiles1.seq.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	DB	ID	Description	
C 1	22.6	43.5	1533	3	US-09-122-210-1	Sequence 1, Appli		
C 2	22.2	42.7	173	3	US-08-542-051-6	Sequence 6, Appli		
C 3	21.2	40.8	7566	2	US-08-232-016-23	Sequence 23, Appli		
C 4	21.2	40.8	7639	2	US-08-232-016-22	Sequence 22, Appli		
C 5	20.6	39.6	6671	1	US-08-280-443-1	Sequence 1, Appli		
C 6	20.6	39.6	6671	1	US-08-457-459-1	Sequence 1, Appli		
C 7	20.6	39.6	6671	1	US-08-555-678-1	Sequence 1, Appli		
C 8	20.6	39.6	6671	5	PCr-US95-02275-1	Sequence 1, Appli		
C 9	20.6	39.6	43280	4	US-08-804-227C-1	Sequence 1, Appli		
C 10	20.2	38.8	340	4	US-08-836-075A-57	Sequence 57, Appli		
C 11	20.2	38.8	2943	1	US-08-042-747A-7	Sequence 7, Appli		
C 12	20.2	38.8	3177	1	US-08-042-747A-4	Sequence 4, Appli		
C 13	20	38.5	269	3	US-09-293-823-1	Sequence 1, Appli		
C 14	20	38.5	840	5	PCr-US91-08177-12	Sequence 12, Appli		
C 15	20	38.5	7323	5	PCr-US91-08177-1	Sequence 1, Appli		
C 16	19.8	38.1	340	4	US-08-836-075A-53	Sequence 53, Appli		
C 17	19.8	38.1	49377	1	US-08-764-233A-1	Sequence 1, Appli		
C 18	19.6	37.7	10095	3	US-08-822-586-45	Sequence 45, Appli		
C 19	19.6	37.7	30001	1	US-08-125-468-1	Sequence 1, Appli		
C 20	19.6	37.7	30001	2	US-08-474-933-1	Sequence 1, Appli		
C 21	19.4	37.3	1001	3	US-08-705-771-6	Sequence 6, Appli		
C 22	19.4	37.3	1499	3	US-08-303-861-17	Sequence 17, Appli		
C 23	19.4	37.3	1788	3	US-08-303-861-1	Sequence 1, Appli		
C 24	19.2	36.9	1893	6	5438136-1	Patent No. 5438126		
C 25	19.2	36.9	1958	1	US-08-115-365-1	Sequence 1, Appli		
C 26	19.2	36.9	1958	1	US-08-586-897-1	Sequence 1, Appli		
C 27	19	36.5	362	4	US-09-060-750-219	Sequence 219, Appli		

C 28	19	36.5	397	4	US-09-060-756-285	Sequence 285, App
C 29	19	36.5	1851	3	US-08-998-416-331	Sequence 331, App
C 30	19	36.5	1833	3	US-08-704-966-3	Sequence 1, Appl
C 31	19	36.5	1833	3	US-08-705-438-3	Sequence 1, Appl
C 32	19	36.5	3758	3	US-08-323-477-1	Sequence 1, Appl
C 33	19	36.5	5996	3	US-09-028-934-33	Sequence 33, Appl
C 34	18.8	36.2	86	1	US-08-447-169A-214	Sequence 214, Appl
C 35	18.8	36.2	1392	4	US-09-163-444-1	Sequence 1, Appl
C 36	18.8	36.2	1608	1	US-07-621-670-2	Sequence 2, Appl
C 37	18.8	36.2	1735	4	US-09-163-444-3	Sequence 3, Appl
C 38	18.8	36.2	2145	1	US-08-849-212-5	Sequence 5, Appl
C 39	18.8	36.2	2703	2	US-08-288-508C-1	Sequence 1, Appl
C 40	18.8	36.2	3252	2	US-08-808-740A-1	Sequence 1, Appl
C 41	18.8	36.2	3252	2	US-08-809-740A-4	Sequence 4, Appl
C 42	18.8	36.2	8438	1	US-07-942-283-1	Sequence 1, Appl
C 43	18.8	36.2	15202	3	US-07-925-635-21	Sequence 21, Appl
C 44	18.6	35.8	340	3	US-08-441-971-9	Sequence 9, Appl
C 45	18.6	35.8	340	4	US-08-442-144A-10	Sequence 10, Appl

ALIGNMENTS

```

RESULT      1
US-09-122-210-1/c
; Sequence 1, Application US/09122210
; Patent No. 6140083
; GENERAL INFORMATION:
; APPLICANT: Haake, David A.
; TITLE OF INVENTION: Leptospiral Outer Membrane Protein
; FILE REFERENCE: 09638/005001
; CURRENT APPLICATION NUMBER: US/09/122,210
; CURRENT FILING DATE: 1998-07-23
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1533
; TYPE: DNA
; ORGANISM: Leptospira sp.
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (78)..(1463)
US-09-122-210-1

```

```
Query Match      43.5%; Score 22.6; DB 3; Length 1533;
Best Local Similarity 68.9%; Pred. NO. 3.3;
Matches 31; Conservative 0; Mismatches 14; Indels 0; Gaps 0;
```

Qy 7 CGTCACTGAGGACCGGClaagcggcccttaaccaaactt 51
||||| |||| |
Dd 1362 CGTTTAGAGCGGAACGGCTAAAGCCGGGGCTTCTCCAGAAGTT 1318
||||| |||| |

```

RESULT      2
US-08-542-051-6/c
; Sequence 6, Application US/08542051F
; Patent No. 6004782
; GENERAL INFORMATION:
; APPLICANT: Daniell, Henry
; APPLICANT: McPherson, David T.
; APPLICANT: Urry, Dan W.
; APPLICANT: Xu, Jie
; TITLE OF INVENTION: Hyperexpression of Bioelastic Polypeptides
; FILE REFERENCE: BERL-018/01US
; CURRENT APPLICATION NUMBER: US/08/542,051F
; CURRENT FILING DATE: 1995-10-13
; EARLIER APPLICATION NUMBER: 08/423,642
; EARLIER FILING DATE: 1995-04-14
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 173

```

; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:synthetic
US-08-542-051-6

Query Match 42.7%; Score 22.2; DB 3; Length 173;
Best Local Similarity 69.8%; Pred. NO. 3.3;
Matches 30; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 2 cacttgcctcactcgagaccgactaaagcgcccttaacc 44
||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 114 CGGTACCAACTCCGGAAGCCTACAGCGGCACACCAACC 72

RESULT 3
US-08-232-016-23/c
; Sequence 23, Application US/08232016
; Patent No. 5952547
; GENERAL INFORMATION:
; APPLICANT: CORNELISSEN, Marc
; APPLICANT: SOETAERT, Piet
; APPLICANT: STAM, Maïke
; APPLICANT: DOCKX, Jan
; APPLICANT: VAN AARSEN, Roel
; TITLE OF INVENTION: MODIFIED GENES AND THEIR EXPRESSION IN
; TITLE OF INVENTION: PLANT CELLS
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,016
; FILING DATE: 03-NOV-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 91402920.2
; FILING DATE: 30-OCT-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 92400820.4
; FILING DATE: 25-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: McGowan, Malcolm K
; REGISTRATION NUMBER: P39,300
; REFERENCE/DOCKET NUMBER: 010830-049
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7566 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: plasmid DNA designated as pPS0212
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1785
; OTHER INFORMATION: /note= "Coding region of a
; OTHER INFORMATION: truncated modified bc2 (cryIab) gene, also designated as the
; OTHER INFORMATION: cryIab6 gene."

; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1793..2026
; OTHER INFORMATION: /note= "3' regulatory sequence
; OTHER INFORMATION: containing the polyadenylation site derived from Agrobacte
; OTHER INFORMATION: T-DNA gene 7."
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 2396..2921
; OTHER INFORMATION: /note= "35S promoter sequence
; OTHER INFORMATION: derived from Cauliflower mosaic virus."
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 2922..3581
; OTHER INFORMATION: /note= "Coding sequence of
; OTHER INFORMATION: chloramphenicol acetyl transferase gene."
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 3582..4407
; OTHER INFORMATION: /note= "3' regulatory sequence
; OTHER INFORMATION: containing the polyadenylation site derived from Agrobacte
; OTHER INFORMATION: T-DNA octopine synthase gene."
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 5600..6457
; OTHER INFORMATION: /note= "Sequence complementary to
; Patent No. 5952547
; OTHER INFORMATION: the coding sequence of the beta-lactamase gene."
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 7071..7566
; OTHER INFORMATION: /note= "Trl' and TR2 promoter
; OTHER INFORMATION: derived from Agrobacterium T-DNA (with modified leader wit
; OTHER INFORMATION: respect to sequence of pJD884 of SEQ ID NO. 22."
US-08-232-016-23

Query Match 40.8%; Score 21.2; DB 2; Length 7566;
Best Local Similarity 76.5%; Pred. No. 15;
Matches 26; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 tcgcttcgctcactcgaggaccggtacaaagccg 34
||||| ||||| ||||| ||||| ||||| |||||
DB 1829 TAGCTTAGCTCATCGGGGATCTGCTAGAGCCGG 1796

RESULT 4
US-08-232-016-22/c
; Sequence 22, Application US/08232016
; Patent No. 5952547
; GENERAL INFORMATION:
; APPLICANT: CORNELISSEN, Marc
; APPLICANT: SOETAERT, Piet
; APPLICANT: STAM, Maïke
; APPLICANT: DOCKX, Jan
; APPLICANT: VAN AARSEN, Roel
; TITLE OF INVENTION: MODIFIED GENES AND THEIR EXPRESSION IN
; TITLE OF INVENTION: PLANT CELLS
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,016

```

/ OTHER INFORMATION: containing the polyadenylation site derived from Agrobacterium
/ OTHER INFORMATION: T-DNA octopine synthase gene."
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: 5684..6541
/ OTHER INFORMATION: /note= "Sequence complementary to
/ Patent No. 5952547
/ OTHER INFORMATION: the coding sequence of the beta-lactamase gene."
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: 7155..7639
/ OTHER INFORMATION: /note= "TR1' and TR2' promoter
/ OTHER INFORMATION: derived from Agrobacterium T-DNA."
/ US-08-2332-016-22

```

```

Query Match      39.6%; Score 20.6; DB 1; Length 6671;
Best Local Similarity 67.4%; Pred. No. 24;
Matches 29; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY       7   cgctcactgcggaccgagctaaacggccgcccttaaccaaacg 49
          ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db       3090 CGTGCGTGCAGGACTGTCAAGAGGGCGCCATCTCCACACG 3048

RESULT
US-08-457-459-1/C
: Sequence 1, Application US/08457459
: Patent No. 5677428
: GENERAL INFORMATION:
: APPLICANT: Nishikura, Kazuko
: TITLE OF INVENTION: RNA Editing Enzyme and Methods of Use
: DATE OF INVENTION: Thereof
: NUMBER OF SEQUENCES: 39
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Howson and Howson
: STREET: Spring House Corporate Cntr., P.O. Box 457
: CITY: Spring House
: STATE: Pennsylvania
: COUNTRY: USA

```



```

? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 08/280,443
? FILING DATE: 25-JUL-1994
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 08/457,459
? FILING DATE: 01-JUN-1995
? ATTORNEY/AGENT INFORMATION:
? NAME: Bak, Mary E.
? REGISTRATION NUMBER: 31,215
? REFERENCE/DOCKET NUMBER: WST49DUSA
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 215-540-9206
? TELEFAX: 215-540-5818
? INFORMATION FOR SEQ ID NO: 1:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 6671 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: double
? TOPOLOGY: unknown
? MOLECULE TYPE: cDNA
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 155..3832
? US-08-555-678-1

Query Match 39.6%; Score 20.6; DB 1; Length 6671;
Best Local Similarity 67.4%; Pred. No. 24;
Matches 29; Conservative 0; Mismatches 14; Indels 0

QY 7 cgctcaactcggggaccgggtctaaacggccctttacccaacq 49
|| || || || || || || || || || || || || || || ||
Db 3090 CGGTGCGTGCAGGACTGTGCAAGAGGCGGCCATCTCCACACG 3048

```

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

```
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6671 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 155..3832
PCT-US95-02275-1

Query Match          39.6%; Score 20.6; DB 5; Length 6671;
Best local Similarity 67.4%; Pred. No. 24;
Matches 29; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 7 cgtcactcgggaccgctaaagccgcgcccttaacaaacy 49
   ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3090 CGTCTCGTCGAGGACTGTCTAAAGAGGCGCCCATCTCCACAG 3048

RESULT 9
US-08-804-227C-1
; Sequence 1, Application US/08804227C
; Patent No. 5876991
; GENERAL INFORMATION:
; APPLICANT: Delhoff, Bradley S.
; APPLICANT: Kuhstoss, Stuart A.
; APPLICANT: Rostock, Paul R., Jr.
; APPLICANT: Sulton, Kimberly L.
; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: THOMAS G. PLANT 1501
; STREET: LILLY CORPORATE CENTER
; CITY: INDIANAPOLIS
; STATE: IN
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: ASCII(DOS) Text only
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/804,227C
; FILING DATE: February 21, 1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Plant, Thomas, G.
; REGISTRATION NUMBER: 35,784
; REFERENCE/DOCKET NUMBER: X-8231
; TELEPHONE: 317-276-2459
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 43280 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 816..14234
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 14351..19945
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 20010..31199
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 31232..36067
```

```
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 36249..41774
US-08-804-227C-1

Query Match          39.6%; Score 20.6; DB 2; Length 43280;
Best local Similarity 74.3%; Pred. No. 13;
Matches 26; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 3 gcttcgctcactcgggaccgctaaagccgcgcc 37
   ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 24347 GCTCGCTCTCTCCGGGCGCGCCGAGGAGGCC 24381

RESULT 10
US-08-836-075A-57
; Sequence 57, Application US/08836075A
; Patent No. 6180768
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; APPLICANT: STUYVER, LIEVEN
; TITLE OF INVENTION: NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES
; TITLE OF INVENTION: AND THEIR USE AS PROPHYLACTIC, THERAPEUTIC AND DIAGNOSTIC
; NUMBER OF SEQUENCES: 207
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ARNOLD, WHITE & DURKEE
; STREET: P.O. BOX 4433
; CITY: HOUSTON
; STATE: TEXAS
; COUNTRY: USA
; ZIP: 77210-4433
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Microsoft Word 6.0 / ASCII text output
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/836,075A
; FILING DATE: 21 Apr 1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP95/04155
; FILING DATE: 23 Oct 1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 94870166.9
; FILING DATE: 21 Oct 1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 95870076.7
; FILING DATE: 28 Jun 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: KAMMERER, PATRICIA A.
; REGISTRATION NUMBER: 29,775
; REFERENCE/DOCKET NUMBER: INNS:004
; INFORMATION FOR SEQ ID NO: 57:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 340 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-836-075A-57
```

```
Query Match          38.8%; Score 20.2; DB 4; Length 340;
Best local Similarity 68.3%; Pred. No. 22;
Matches 28; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 6 tcgctcactcgggaccgctaaagccgcgcccttaacaa 46
   ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 89 TCGCTCACCAGCGGCTGTATATCGGGGTCCCTTAACCAA 129
```

```
RESULT 11
US-08-042-747A-7
; Sequence 7, Application US/08042747A
; Patent No. 5487969
; GENERAL INFORMATION:
; APPLICANT: Eberle, Richard
; APPLICANT: Black, Darla
; APPLICANT: Scinicariello, Franco
; APPLICANT: Hilliard, Julia K.
; TITLE OF INVENTION: Cloning and Amplification of Monkey B
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cox & Smith Incorporated
; STREET: 112 East Pecan Street, Suite 2000
; CITY: San Antonio
; STATE: Texas
; COUNTRY: USA
; ZIP: 78205
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/042.747A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Haymond, W. Bradley
; REGISTRATION NUMBER: 35186
; REFERENCE/DOCKET NUMBER: S-0072.179
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 210-554-5500
; TELEFAX: 210-226-8395
; TELEX: 767609
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2943 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 87..2744
; US-08-042-747A-7

Query Match 38.8%; Score 20.2; DB 1; Length 2943;
Best Local Similarity 75.8%; Pred. No. 31;
Matches 25; Conservative 0; Mismatches 8; Indels 0; Caps 0;

Qy 7 cgtcactgcgggaccgctaagccggccct 39
||| | | ||||| |||| ||||| |||||
Db 1811 CGCGCGCCCGGGAGCTGCTACACGCCGCCCT 1843

RESULT 12
US-08-042-747A-4
; Sequence 4, Application US/08042747A
; Patent No. 5487969
; GENERAL INFORMATION:
; APPLICANT: Eberle, Richard
; APPLICANT: Black, Darla
; APPLICANT: Scinicariello, Franco
; APPLICANT: Hilliard, Julia K.
; TITLE OF INVENTION: Cloning and Amplification of Monkey B
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
```

```
; ADDRESSEE: Cox & Smith Incorporated
; STREET: 112 East Pecan Street, Suite 2000
; CITY: San Antonio
; STATE: Texas
; COUNTRY: USA
; ZIP: 78205
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/042.747A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Haymond, W. Bradley
; REGISTRATION NUMBER: 35186
; REFERENCE/DOCKET NUMBER: S-0072.179
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 210-554-5500
; TELEFAX: 210-226-8395
; TELEX: 767609
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3177 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 269..2941
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..249
; US-08-042-747A-4

Query Match 38.8%; Score 20.2; DB 1; Length 3177;
Best Local Similarity 75.8%; Pred. No. 31;
Matches 25; Conservative 0; Mismatches 8; Indels 0; Caps 0;

Qy 7 cgtcactgcgggaccgctaagccggccct 39
||| | | ||||| |||| ||||| |||||
Db 1999 CGCGCGCCCGGGAGCTGCTACACGCCGCCCT 2031

RESULT 13
US-09-293-823-1/c
; Sequence 1, Application US/09293823
; Patent No. 6126955
; GENERAL INFORMATION:
; APPLICANT: William D. Huse
; TITLE OF INVENTION: A NEW METHOD FOR TAPPING THE
; FILE REFERENCE: 62710024US06
; CURRENT APPLICATION NUMBER: US/09/293,823
; CURRENT FILING DATE: 1999-04-16
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 269
; TYPE: DNA
; ORGANISM: Synthetic DNA
; US-09-293-823-1

Query Match 38.5%; Score 20; DB 3; Length 269;
Best Local Similarity 72.2%; Pred. No. 25;
Matches 26; Conservative 0; Mismatches 10; Indels 0; Caps 0;

Qy 8 gctcactgcgggaccgctaagccggcccttaac 43
```


Best Local Similarity 65.9%; Pred. No. 42;
Matches 29; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Qy 5 ttgcctactgaggaccggctaaagccggcccttaaccaaac 48
|| | | | | | | | | | | | | | | | | | | | |
Db 3743 TTACATTAGTGTGGAAATGCTAAAGCCAGCCACATCCAAC 3786

Search completed: August 2, 2001, 18:40:43
Job time: 6333 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw method

Run on: August 2, 2001, 17:44:12 Search time 2788.81 Seconds
(without alignments)
275.528 Million cell updates/sec

Title: US-09-701-626a-79
Perfect score: 52
Sequence: 1 tcgcttgctcactgcggga.....ggcccttaaccaaacgtta 52

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 14155048 seqs, 7388405095 residues
Total number of hits satisfying chosen parameters: 28310096

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Parents_NA_Main.*
1: /cgnl_7/ptodata/1/pna/PCTUS_COMB.seq.*
2: /cgnl_7/ptodata/1/pna/US06_COMB.seq.*
3: /cgnl_7/ptodata/1/pna/US07_COMB.seq.*
4: /cgnl_7/ptodata/1/pna/US080_COMB.seq.*
5: /cgnl_7/ptodata/1/pna/US081_COMB.seq.*
6: /cgnl_7/ptodata/1/pna/US082_COMB.seq.*
7: /cgnl_7/ptodata/1/pna/US083_COMB.seq.*
8: /cgnl_7/ptodata/1/pna/US084_COMB.seq.*
9: /cgnl_7/ptodata/1/pna/US085_COMB.seq.*
10: /cgnl_7/ptodata/1/pna/US086_COMB.seq.*
11: /cgnl_7/ptodata/1/pna/US087_COMB.seq.*
12: /cgnl_7/ptodata/1/pna/US088_COMB.seq.*
13: /cgnl_7/ptodata/1/pna/US089_COMB.seq.*
14: /cgnl_7/ptodata/1/pna/US090_COMB.seq.*
15: /cgnl_7/ptodata/1/pna/US091_COMB.seq.*
16: /cgnl_7/ptodata/1/pna/US092_COMB.seq.*
17: /cgnl_7/ptodata/1/pna/US093_COMB.seq.*
18: /cgnl_7/ptodata/1/pna/US094_COMB.seq.*
19: /cgnl_7/ptodata/1/pna/US095A_COMB.seq.*
20: /cgnl_7/ptodata/1/pna/US095B_COMB.seq.*
21: /cgnl_7/ptodata/1/pna/US095C_COMB.seq.*
22: /cgnl_7/ptodata/1/pna/US095D_COMB.seq.*
23: /cgnl_7/ptodata/1/pna/US096A_COMB.seq.*
24: /cgnl_7/ptodata/1/pna/US096B_COMB.seq.*
25: /cgnl_7/ptodata/1/pna/US096C_COMB.seq.*
26: /cgnl_7/ptodata/1/pna/US096D_COMB.seq.*
27: /cgnl_7/ptodata/1/pna/US096E_COMB.seq.*
28: /cgnl_7/ptodata/1/pna/US097A_COMB.seq.*
29: /cgnl_7/ptodata/1/pna/US097B_COMB.seq.*
30: /cgnl_7/ptodata/1/pna/US097C_COMB.seq.*
31: /cgnl_7/ptodata/1/pna/US098_COMB.seq.*
32: /cgnl_7/ptodata/1/pna/US6000_COMB.seq.*
33: /cgnl_7/ptodata/1/pna/US6001_COMB.seq.*
34: /cgnl_7/ptodata/1/pna/US6002_COMB.seq.*
35: /cgnl_7/ptodata/1/pna/US6003_COMB.seq.*
36: /cgnl_7/ptodata/1/pna/US6004_COMB.seq.*
37: /cgnl_7/ptodata/1/pna/US6005_COMB.seq.*
38: /cgnl_7/ptodata/1/pna/US6006_COMB.seq.*
39: /cgnl_7/ptodata/1/pna/US6007_COMB.seq.*
40: /cgnl_7/ptodata/1/pna/US6008_COMB.seq.*
41: /cgnl_7/ptodata/1/pna/US6009_COMB.seq.*
42: /cgnl_7/ptodata/1/pna/US6010_COMB.seq.*
43: /cgnl_7/ptodata/1/pna/US6011_COMB.seq.*

44: /cgnl_7/ptodata/1/pna/US6012_COMB.seq.*
45: /cgnl_7/ptodata/1/pna/US6013_COMB.seq.*
46: /cgnl_7/ptodata/1/pna/US6014_COMB.seq.*
47: /cgnl_7/ptodata/1/pna/US6015_COMB.seq.*
48: /cgnl_7/ptodata/1/pna/US6016_COMB.seq.*
49: /cgnl_7/ptodata/1/pna/US6017_COMB.seq.*
50: /cgnl_7/ptodata/1/pna/US6018_COMB.seq.*
51: /cgnl_7/ptodata/1/pna/US6019_COMB.seq.*
52: /cgnl_7/ptodata/1/pna/US6020_COMB.seq.*
53: /cgnl_7/ptodata/1/pna/US6021_COMB.seq.*
54: /cgnl_7/ptodata/1/pna/US6022_COMB.seq.*
55: /cgnl_7/ptodata/1/pna/US6023_COMB.seq.*
56: /cgnl_7/ptodata/1/pna/US6024_COMB.seq.*
57: /cgnl_7/ptodata/1/pna/US6025_COMB.seq.*
58: /cgnl_7/ptodata/1/pna/US6026_COMB.seq.*
59: /cgnl_7/ptodata/1/pna/US6027_COMB.seq.*
60: /cgnl_7/ptodata/1/pna/US6028_COMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	52	100.0	52	1	PCT-US99-13295-79	Sequence 79, Appl
2	52	100.0	52	28	US-09-701-626A-79	Sequence 79, Appl
3	52	100.0	78	1	PCT-US99-13295-23	Sequence 23, Appl
4	52	100.0	78	1	PCT-US99-13295-38	Sequence 38, Appl
5	52	100.0	78	1	PCT-US99-13295-47	Sequence 47, Appl
6	52	100.0	78	1	PCT-US99-13295-49	Sequence 49, Appl
7	52	100.0	78	1	PCT-US99-13295-51	Sequence 51, Appl
8	52	100.0	78	1	PCT-US99-13295-59	Sequence 59, Appl
9	52	100.0	78	28	US-09-701-626A-23	Sequence 23, Appl
10	52	100.0	78	28	US-09-701-626A-38	Sequence 38, Appl
11	52	100.0	78	28	US-09-701-626A-47	Sequence 47, Appl
12	52	100.0	78	28	US-09-701-626A-49	Sequence 49, Appl
13	52	100.0	78	28	US-09-701-626A-51	Sequence 51, Appl
14	52	100.0	78	28	US-09-701-626A-59	Sequence 59, Appl
15	52	100.0	7300	1	PCT-US99-13295-3	Sequence 3, Appl
16	52	100.0	7300	28	US-09-701-626A-2	Sequence 2, Appl
17	52	100.0	7300	28	US-09-701-626A-3	Sequence 3, Appl
18	52	100.0	12956	28	US-09-701-626A-4	Sequence 4, Appl
19	52	100.0	14143	1	PCT-US99-13295-1	Sequence 1, Appl
20	52	100.0	14143	1	PCT-US99-13295-2	Sequence 2, Appl
21	52	100.0	14143	28	US-09-701-626A-1	Sequence 1, Appl
22	52	100.0	42143	1	PCT-US99-13295-4	Sequence 4, Appl
23	50.4	96.9	78	1	PCT-US99-13295-67	Sequence 67, Appl
24	50.4	96.9	78	28	US-09-701-626A-67	Sequence 67, Appl
25	49.4	95.0	79	1	PCT-US99-13295-52	Sequence 52, Appl
26	49.4	95.0	79	28	US-09-701-626A-52	Sequence 52, Appl
27	40	76.9	77	1	PCT-US99-13295-11	Sequence 11, Appl
28	40	76.9	77	1	PCT-US99-13295-19	Sequence 19, Appl
29	40	76.9	77	1	PCT-US99-13295-22	Sequence 22, Appl
30	40	76.9	77	1	PCT-US99-13295-24	Sequence 24, Appl
31	40	76.9	77	1	PCT-US99-13295-41	Sequence 41, Appl
32	40	76.9	77	1	PCT-US99-13295-55	Sequence 55, Appl
33	40	76.9	77	1	PCT-US99-13295-63	Sequence 63, Appl
34	40	76.9	77	1	PCT-US99-13295-65	Sequence 65, Appl
35	40	76.9	77	1	PCT-US99-13295-68	Sequence 68, Appl
36	40	76.9	77	1	PCT-US99-13295-77	Sequence 77, Appl
37	40	76.9	77	1	PCT-US99-13295-78	Sequence 78, Appl
38	40	76.9	77	1	PCT-US99-13295-119	Sequence 119, Appl
39	40	76.9	77	28	US-09-701-626A-11	Sequence 11, Appl
40	40	76.9	77	28	US-09-701-626A-19	Sequence 19, Appl
41	40	76.9	77	28	US-09-701-626A-22	Sequence 22, Appl
42	40	76.9	77	28	US-09-701-626A-24	Sequence 24, Appl
43	40	76.9	77	28	US-09-701-626A-41	Sequence 41, Appl
44	40	76.9	77	28	US-09-701-626A-44	Sequence 44, Appl
45	40	76.9	77	28	US-09-701-626A-55	Sequence 55, Appl

ALIGNMENTS

RESULT 1
PCT-US99-13295-79
; Sequence 79, Application PC/TUS9913295
; GENERAL INFORMATION:
; APPLICANT: New England Biolabs, Inc.
; APPLICANT: Vaisvila, Romualdas
; APPLICANT: Morgan, Richard D.
; APPLICANT: Raleigh, Elisabeth
; TITLE OF INVENTION: Restriction Enzyme Gene Discovery Method
; FILE REFERENCE: Gene Discovery Method
; CURRENT APPLICATION NUMBER: PCT/US99/13295
; CURRENT FILING DATE: 1999-06-11
; EARLIER APPLICATION NUMBER: 60/089,086
; EARLIER FILING DATE: 1998-06-12
; EARLIER APPLICATION NUMBER: 60/089,101
; EARLIER FILING DATE: 1998-06-12
; NUMBER OF SEQ ID NOS: 130
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 79
; LENGTH: 52
; TYPE: DNA
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: Synthetic
; OTHER INFORMATION: Oligonucleotide based on Pseudomonas Alcaligenes
; OTHER INFORMATION: NEB#585 (ATCC 55044)
PCT-US99-13295-79

Query Match 100.0%; Score 52; DB 1; Length 52;
Best Local Similarity 100.0%; Pred. No. 8.2e-10;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tcgcttcgctcactgcgggaccggtctaaagcggcccttaacaaacgtta 52
|||||
Db 1 tcgcttcgctcactgcgggaccggtctaaagcggcccttaacaaacgtta 52

RESULT 2
US-09-701-626A-79
; Sequence 79, Application US/09701626A
; GENERAL INFORMATION:
; APPLICANT: Vaisvila, Romualdas
; APPLICANT: Morgan, Richard D.
; APPLICANT: Raleigh, Elisabeth
; TITLE OF INVENTION: Restriction Enzyme Gene Discovery Method
; FILE REFERENCE: NEB-165PUS
; CURRENT APPLICATION NUMBER: US/09/701,626A
; CURRENT FILING DATE: 2001-04-16
; PRIOR APPLICATION NUMBER: 60/089,086
; PRIOR FILING DATE: 1998-06-12
; PRIOR APPLICATION NUMBER: 60/089,101
; PRIOR FILING DATE: 1998-06-12
; PRIOR APPLICATION NUMBER: PCT/US99/13295
; PRIOR FILING DATE: 1999-06-11
; NUMBER OF SEQ ID NOS: 130
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 79
; LENGTH: 52
; TYPE: DNA
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: Synthetic
; OTHER INFORMATION: Oligonucleotide based on Pseudomonas Alcaligenes
; OTHER INFORMATION: NEB#585 (ATCC 55044)
US-09-701-626A-79

Query Match 100.0%; Score 52; DB 28; Length 52;

Best Local Similarity 100.0%; Pred. No. 8.2e-10;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tcgcttcgctcactgcgggaccggtctaaagcggcccttaacaaacgtta 52
|||||
Db 1 tcgcttcgctcactgcgggaccggtctaaagcggcccttaacaaacgtta 52

RESULT 3
PCT-US99-13295-23
; Sequence 23, Application PC/TUS9913295
; GENERAL INFORMATION:
; APPLICANT: New England Biolabs, Inc.
; APPLICANT: Vaisvila, Romualdas
; APPLICANT: Morgan, Richard D.
; APPLICANT: Raleigh, Elisabeth
; TITLE OF INVENTION: Restriction Enzyme Gene Discovery Method
; FILE REFERENCE: Gene Discovery Method
; CURRENT APPLICATION NUMBER: PCT/US99/13295
; CURRENT FILING DATE: 1999-06-11
; EARLIER APPLICATION NUMBER: 60/089,086
; EARLIER FILING DATE: 1998-06-12
; EARLIER APPLICATION NUMBER: 60/089,101
; EARLIER FILING DATE: 1998-06-12
; NUMBER OF SEQ ID NOS: 130
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 23
; LENGTH: 78
; TYPE: DNA
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: Genomic DNA of
; OTHER INFORMATION: Pseudomonas Alcaligenes NEB #585 (ATCC 55044)
PCT-US99-13295-23

Query Match 100.0%; Score 52; DB 1; Length 78;
Best Local Similarity 100.0%; Pred. No. 8.5e-10;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tcgcttcgctcactgcgggaccggtctaaagcggcccttaacaaacgtta 52
|||||
Db 24 tcgcttcgctcactgcgggaccggtctaaagcggcccttaacaaacgtta 75

RESULT 4
PCT-US99-13295-38
; Sequence 38, Application PC/TUS9913295
; GENERAL INFORMATION:
; APPLICANT: New England Biolabs, Inc.
; APPLICANT: Vaisvila, Romualdas
; APPLICANT: Morgan, Richard D.
; APPLICANT: Raleigh, Elisabeth
; TITLE OF INVENTION: Restriction Enzyme Gene Discovery Method
; FILE REFERENCE: Gene Discovery Method
; CURRENT APPLICATION NUMBER: PCT/US99/13295
; CURRENT FILING DATE: 1999-06-11
; EARLIER APPLICATION NUMBER: 60/089,086
; EARLIER FILING DATE: 1998-06-12
; EARLIER APPLICATION NUMBER: 60/089,101
; EARLIER FILING DATE: 1998-06-12
; NUMBER OF SEQ ID NOS: 130
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 38
; LENGTH: 78
; TYPE: DNA
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: Genomic DNA of
; OTHER INFORMATION: Pseudomonas Alcaligenes NEB #585 (ATCC 55044)
PCT-US99-13295-38

Query Match 100.0%; Score 52; DB 1; Length 78;
Best Local Similarity 100.0%; Pred. No. 8.5e-10;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcgcttcgctcactcgggaccggtctaaagccggcccttaacccaaacgtta 52
|||||
Db 24 tcgcttcgctcactcgggaccggtctaaagccggcccttaacccaaacgtta 75
|||||

RESULT 5

PCT-US99-13295-47
; Sequence 47, Application PC/TUS9913295
; GENERAL INFORMATION:
; APPLICANT: New England Biolabs, Inc.
; APPLICANT: Vaisvila, Romualdas
; APPLICANT: Morgan, Richard D.
; APPLICANT: Raleigh, Elisabeth
; TITLE OF INVENTION: Restriction Enzyme Gene Discovery Method
; FILE REFERENCE: Gene Discovery Method
; CURRENT APPLICATION NUMBER: PCT/US99/13295
; EARLIER FILING DATE: 1999-06-11
; EARLIER APPLICATION NUMBER: 60/089,086
; EARLIER FILING DATE: 1998-06-12
; EARLIER APPLICATION NUMBER: 60/089,101
; EARLIER FILING DATE: 1998-06-12
; NUMBER OF SEQ ID NOS: 130
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 47
; LENGTH: 78
; TYPE: DNA
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: Genomic DNA of
; OTHER INFORMATION: Pseudomonas Alcaligenes NEB #585 (ATCC 55044)
PCT-US99-13295-47

Query Match 100.0%; Score 52; DB 1; Length 78;
Best Local Similarity 100.0%; Pred. No. 8.5e-10;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcgcttcgctcactcgggaccggtctaaagccggcccttaacccaaacgtta 52
|||||
Db 24 tcgcttcgctcactcgggaccggtctaaagccggcccttaacccaaacgtta 75
|||||

RESULT 6

PCT-US99-13295-49
; Sequence 49, Application PC/TUS9913295
; GENERAL INFORMATION:
; APPLICANT: New England Biolabs, Inc.
; APPLICANT: Vaisvila, Romualdas
; APPLICANT: Morgan, Richard D.
; APPLICANT: Raleigh, Elisabeth
; TITLE OF INVENTION: Restriction Enzyme Gene Discovery Method
; FILE REFERENCE: Gene Discovery Method
; CURRENT APPLICATION NUMBER: PCT/US99/13295
; EARLIER FILING DATE: 1999-06-11
; EARLIER APPLICATION NUMBER: 60/089,086
; EARLIER FILING DATE: 1998-06-12
; EARLIER APPLICATION NUMBER: 60/089,101
; EARLIER FILING DATE: 1998-06-12
; NUMBER OF SEQ ID NOS: 130
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 49
; LENGTH: 78
; TYPE: DNA
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: Genomic DNA of
; OTHER INFORMATION: Pseudomonas Alcaligenes NEB #585 (ATCC 55044)
PCT-US99-13295-49

Query Match 100.0%; Score 52; DB 1; Length 78;
Best Local Similarity 100.0%; Pred. No. 8.5e-10;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcgcttcgctcactcgggaccggtctaaagccggcccttaacccaaacgtta 52
|||||
Db 24 tcgcttcgctcactcgggaccggtctaaagccggcccttaacccaaacgtta 75
|||||

RESULT 7

PCT-US99-13295-51
; Sequence 51, Application PC/TUS9913295
; GENERAL INFORMATION:
; APPLICANT: New England Biolabs, Inc.
; APPLICANT: Vaisvila, Romualdas
; APPLICANT: Morgan, Richard D.
; APPLICANT: Raleigh, Elisabeth
; TITLE OF INVENTION: Restriction Enzyme Gene Discovery Method
; FILE REFERENCE: Gene Discovery Method
; CURRENT APPLICATION NUMBER: PCT/US99/13295
; EARLIER FILING DATE: 1999-06-11
; EARLIER APPLICATION NUMBER: 60/089,086
; EARLIER FILING DATE: 1998-06-12
; EARLIER APPLICATION NUMBER: 60/089,101
; EARLIER FILING DATE: 1998-06-12
; NUMBER OF SEQ ID NOS: 130
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 51
; LENGTH: 78
; TYPE: DNA
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: Genomic DNA of
; OTHER INFORMATION: Pseudomonas Alcaligenes NEB #585 (ATCC 55044)
PCT-US99-13295-51

Query Match 100.0%; Score 52; DB 1; Length 78;
Best Local Similarity 100.0%; Pred. No. 8.5e-10;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcgcttcgctcactcgggaccggtctaaagccggcccttaacccaaacgtta 52
|||||
Db 24 tcgcttcgctcactcgggaccggtctaaagccggcccttaacccaaacgtta 75
|||||

RESULT 8

PCT-US99-13295-59
; Sequence 59, Application PC/TUS9913295
; GENERAL INFORMATION:
; APPLICANT: New England Biolabs, Inc.
; APPLICANT: Vaisvila, Romualdas
; APPLICANT: Morgan, Richard D.
; APPLICANT: Raleigh, Elisabeth
; TITLE OF INVENTION: Restriction Enzyme Gene Discovery Method
; FILE REFERENCE: Gene Discovery Method
; CURRENT APPLICATION NUMBER: PCT/US99/13295
; EARLIER FILING DATE: 1999-06-11
; EARLIER APPLICATION NUMBER: 60/089,086
; EARLIER FILING DATE: 1998-06-12
; EARLIER APPLICATION NUMBER: 60/089,101
; EARLIER FILING DATE: 1998-06-12
; NUMBER OF SEQ ID NOS: 130
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 59
; LENGTH: 78
; TYPE: DNA
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: Genomic DNA of
; OTHER INFORMATION: Pseudomonas Alcaligenes NEB #585 (ATCC 55044)
PCT-US99-13295-59


```
; OTHER INFORMATION: Description of Unknown Organism: Genomic DNA of
; OTHER INFORMATION: Pseudomonas Alcaligenes NEB #585 (ATCC 55044)
US-09-701-626A-38

Query Match          100.0%; Score 52; DB 1; Length 78;
Best Local Similarity 100.0%; Pred. No. 8.5e-10;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcgcttcgctcaactgcgggaccggtctaaagccggcccttaacccaacgtta 52
   |||||||
Db 24 tcgcttcgctcaactgcgggaccggtctaaagccggcccttaacccaacgtta 75

RESULT 9
US-09-701-626A-23
; Sequence 23, Application US/09701626A
; GENERAL INFORMATION:
; APPLICANT: Vaisvila, Romualdas
; APPLICANT: Morgan, Richard D.
; APPLICANT: Raleigh, Elisabeth
; TITLE OF INVENTION: Restriction Enzyme Gene Discovery Method
; FILE REFERENCE: NEB-165PUS
; CURRENT APPLICATION NUMBER: US/09/701,626A
; CURRENT FILING DATE: 2001-04-16
; PRIOR FILING DATE: 1998-06-12
; PRIOR APPLICATION NUMBER: 60/089,101
; PRIOR FILING DATE: 1998-06-12
; PRIOR APPLICATION NUMBER: PCT/US99/13295
; PRIOR FILING DATE: 1999-06-11
; NUMBER OF SEQ ID NOS: 130
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 23
; LENGTH: 78
; TYPE: DNA
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: Genomic DNA of
; OTHER INFORMATION: Pseudomonas Alcaligenes NEB #585 (ATCC 55044)
US-09-701-626A-23

Query Match          100.0%; Score 52; DB 28; Length 78;
Best Local Similarity 100.0%; Pred. No. 8.5e-10;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcgcttcgctcaactgcgggaccggtctaaagccggcccttaacccaacgtta 52
   |||||||
Db 24 tcgcttcgctcaactgcgggaccggtctaaagccggcccttaacccaacgtta 75

RESULT 10
US-09-701-626A-38
; Sequence 38, Application US/09701626A
; GENERAL INFORMATION:
; APPLICANT: Vaisvila, Romualdas
; APPLICANT: Morgan, Richard D.
; APPLICANT: Raleigh, Elisabeth
; TITLE OF INVENTION: Restriction Enzyme Gene Discovery Method
; FILE REFERENCE: NEB-165PUS
; CURRENT APPLICATION NUMBER: US/09/701,626A
; CURRENT FILING DATE: 2001-04-16
; PRIOR FILING DATE: 1998-06-12
; PRIOR APPLICATION NUMBER: 60/089,101
; PRIOR FILING DATE: 1998-06-12
; PRIOR APPLICATION NUMBER: PCT/US99/13295
; PRIOR FILING DATE: 1999-06-11
; NUMBER OF SEQ ID NOS: 130
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 38
; LENGTH: 78
; TYPE: DNA
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: Genomic DNA of
; OTHER INFORMATION: Pseudomonas Alcaligenes NEB #585 (ATCC 55044)
US-09-701-626A-47

Query Match          100.0%; Score 52; DB 28; Length 78;
Best Local Similarity 100.0%; Pred. No. 8.5e-10;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcgcttcgctcaactgcgggaccggtctaaagccggcccttaacccaacgtta 52
   |||||||
Db 24 tcgcttcgctcaactgcgggaccggtctaaagccggcccttaacccaacgtta 75

RESULT 11
US-09-701-626A-47
; Sequence 47, Application US/09701626A
; GENERAL INFORMATION:
; APPLICANT: Vaisvila, Romualdas
; APPLICANT: Morgan, Richard D.
; APPLICANT: Raleigh, Elisabeth
; TITLE OF INVENTION: Restriction Enzyme Gene Discovery Method
; FILE REFERENCE: NEB-165PUS
; CURRENT APPLICATION NUMBER: US/09/701,626A
; CURRENT FILING DATE: 2001-04-16
; PRIOR FILING DATE: 1998-06-12
; PRIOR APPLICATION NUMBER: 60/089,101
; PRIOR FILING DATE: 1998-06-12
; PRIOR APPLICATION NUMBER: PCT/US99/13295
; PRIOR FILING DATE: 1999-06-11
; NUMBER OF SEQ ID NOS: 130
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 47
; LENGTH: 78
; TYPE: DNA
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: Genomic DNA of
; OTHER INFORMATION: Pseudomonas Alcaligenes NEB #585 (ATCC 55044)
US-09-701-626A-47

Query Match          100.0%; Score 52; DB 28; Length 78;
Best Local Similarity 100.0%; Pred. No. 8.5e-10;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcgcttcgctcaactgcgggaccggtctaaagccggcccttaacccaacgtta 52
   |||||||
Db 24 tcgcttcgctcaactgcgggaccggtctaaagccggcccttaacccaacgtta 75

RESULT 12
US-09-701-626A-49
; Sequence 49, Application US/09701626A
; GENERAL INFORMATION:
; APPLICANT: Vaisvila, Romualdas
; APPLICANT: Morgan, Richard D.
; APPLICANT: Raleigh, Elisabeth
; TITLE OF INVENTION: Restriction Enzyme Gene Discovery Method
; FILE REFERENCE: NEB-165PUS
; CURRENT APPLICATION NUMBER: US/09/701,626A
; CURRENT FILING DATE: 2001-04-16
; PRIOR FILING DATE: 1998-06-12
; PRIOR APPLICATION NUMBER: 60/089,101
; PRIOR FILING DATE: 1998-06-12
; PRIOR APPLICATION NUMBER: PCT/US99/13295
; PRIOR FILING DATE: 1999-06-11
; NUMBER OF SEQ ID NOS: 130
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 49
; LENGTH: 78
; TYPE: DNA
; ORGANISM: Unknown
; FEATURE:
```

; TYPE: DNA
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: Genomic DNA of
; OTHER INFORMATION: Pseudomonas Alcaligenes NEB #585 (ATCC 55044)
US-09-701-626A-49

Query Match 100.0%; Score 52; DB 28; Length 78;
Best Local Similarity 100.0%; Pred. No. 8.5e-10;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcgcttcgctcactgcgggaccggtctaaagccggcccttaacccaaacgtta 52
|||||
Db 24 tcgcttcgctcactgcgggaccggtctaaagccggcccttaacccaaacgtta 75

RESULT 13

US-09-701-626A-51

; Sequence 51, Application US/09701626A
; GENERAL INFORMATION:
; APPLICANT: Vaisvila, Romualdas
; APPLICANT: Morgan, Richard D.
; APPLICANT: Raleigh, Elisabeth
; TITLE OF INVENTION: Restriction Enzyme Gene Discovery Method
; FILE REFERENCE: NEB-165PUS
; CURRENT APPLICATION NUMBER: US/09/701,626A
; PRIOR FILING DATE: 2001-04-16
; PRIOR APPLICATION NUMBER: 60/089,086
; PRIOR FILING DATE: 1998-06-12
; PRIOR APPLICATION NUMBER: 60/089,101
; PRIOR FILING DATE: 1998-06-12
; PRIOR APPLICATION NUMBER: PCT/US99/13295
; PRIOR FILING DATE: 1999-06-11
; NUMBER OF SEQ ID NOS: 130
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 51
; LENGTH: 78
; TYPE: DNA
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: Genomic DNA of
; OTHER INFORMATION: Pseudomonas Alcaligenes NEB #585 (ATCC 55044)
US-09-701-626A-51

Query Match 100.0%; Score 52; DB 28; Length 78;
Best Local Similarity 100.0%; Pred. No. 8.5e-10;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcgcttcgctcactgcgggaccggtctaaagccggcccttaacccaaacgtta 52
|||||
Db 24 tcgcttcgctcactgcgggaccggtctaaagccggcccttaacccaaacgtta 75

RESULT 14

US-09-701-626A-59

; Sequence 59, Application US/09701626A
; GENERAL INFORMATION:
; APPLICANT: Vaisvila, Romualdas
; APPLICANT: Morgan, Richard D.
; APPLICANT: Raleigh, Elisabeth
; TITLE OF INVENTION: Restriction Enzyme Gene Discovery Method
; FILE REFERENCE: NEB-165PUS
; CURRENT APPLICATION NUMBER: US/09/701,626A
; CURRENT FILING DATE: 2001-04-16
; PRIOR APPLICATION NUMBER: 60/089,086
; PRIOR FILING DATE: 1998-06-12
; PRIOR APPLICATION NUMBER: 60/089,101
; PRIOR FILING DATE: 1998-06-12
; PRIOR APPLICATION NUMBER: PCT/US99/13295
; PRIOR FILING DATE: 1999-06-11
; NUMBER OF SEQ ID NOS: 130

; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 59
; LENGTH: 78
; TYPE: DNA
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: Genomic DNA of
; OTHER INFORMATION: Pseudomonas Alcaligenes NEB #585 (ATCC 55044)
US-09-701-626A-59

Query Match 100.0%; Score 52; DB 28; Length 78;
Best Local Similarity 100.0%; Pred. No. 8.5e-10;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcgcttcgctcactgcgggaccggtctaaagccggcccttaacccaaacgtta 52
|||||
Db 24 tcgcttcgctcactgcgggaccggtctaaagccggcccttaacccaaacgtta 75

RESULT 15

PCT-US99-13295-3

; Sequence 3, Application PC/TUS9913295
; GENERAL INFORMATION:
; APPLICANT: New England Biolabs, Inc.
; APPLICANT: Vaisvila, Romualdas
; APPLICANT: Morgan, Richard D.
; APPLICANT: Raleigh, Elisabeth
; TITLE OF INVENTION: Restriction Enzyme Gene Discovery Method
; FILE REFERENCE: Gene Discovery Method
; CURRENT APPLICATION NUMBER: PCT/US99/13295
; CURRENT FILING DATE: 1999-06-11
; EARLIER APPLICATION NUMBER: 60/089,086
; EARLIER FILING DATE: 1998-06-12
; EARLIER APPLICATION NUMBER: 60/089,101
; EARLIER FILING DATE: 1998-06-12
; NUMBER OF SEQ ID NOS: 130
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 7300
; TYPE: DNA
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: Genomic DNA of
; OTHER INFORMATION: Pseudomonas Alcaligenes NEB #585 (ATCC 55044)
PCT-US99-13295-3

Query Match 100.0%; Score 52; DB 1; Length 7300;
Best Local Similarity 100.0%; Pred. No. 1.3e-09;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcgcttcgctcactgcgggaccggtctaaagccggcccttaacccaaacgtta 52
|||||
Db 3300 tcgcttcgctcactgcgggaccggtctaaagccggcccttaacccaaacgtta 3351

Search completed: August 2, 2001, 17:44:12
Job time: 3062 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 2, 2001, 16:54:36 ; Search time 49.77 Seconds
(without alignments)
730.055 Million cell updates/sec

Title: US-09-701-626A-79
Perfect score: 52
Sequence: 1 tcgttcgctcactgcggga.....ggcccttaacaaacgtta 52

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 327855 seqs, 34937361 residues

Total number of hits satisfying chosen parameters: 655710

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending Patents, NA, New.*
1: /cgnl_7/ptodata/2/pna/PCT_NEW_COMB.seq.*
2: /cgnl_7/ptodata/2/pna/US06_NEW_COMB.seq.*
3: /cgnl_7/ptodata/2/pna/US07_NEW_COMB.seq.*
4: /cgnl_7/ptodata/2/pna/US08_NEW_COMB.seq.*
5: /cgnl_7/ptodata/2/pna/US09_NEW_COMB.seq.*
6: /cgnl_7/ptodata/2/pna/US60_NEW_COMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	23.4	45.0	422	5	US-09-804-730-10706
2	22.4	43.1	388	5	US-09-904-703-1102
3	22.4	43.1	531	5	US-09-904-703-14754
4	22.4	43.1	1409	5	US-09-758-474-201
5	22.2	42.7	837	5	US-09-803-110-6363
6	22.2	42.7	1395	5	US-09-803-110-4177
7	22.2	42.7	3157	5	US-09-764-902-493
8	22.2	42.7	8416	6	US-60-278-232-8896
9	22.2	42.7	12051	5	US-09-764-902-2125
10	22.2	42.7	18416	5	US-09-764-902-2124
11	22.2	42.7	260744	5	US-09-803-110-211
12	22.2	42.7	506113	5	US-09-803-110-218
13	21.6	41.5	1029	5	US-09-803-110-3210
14	21.6	41.5	1145	5	US-09-758-457-139
15	21.6	41.5	4202	6	US-60-278-232-2701
16	21.6	41.5	78844	5	US-09-803-736-726
17	21.6	41.5	184668	5	US-09-803-110-207
18	21.2	40.8	3207	5	US-09-803-110-3990
19	21.2	40.8	244769	5	US-09-803-110-210
20	21.2	40.4	4840	6	US-60-278-258-11152
21	20.8	40.0	315	5	US-09-757-032-605
22	20.8	40.0	86014	5	US-09-803-736-1009
23	20.6	39.6	389	5	US-09-804-730-13213
24	20.6	39.6	393	5	US-09-803-110-6830
25	20.6	39.6	398	5	US-09-904-703-6327
26	20.6	39.6	414	5	US-09-764-902-222
27	20.6	39.6	2641	5	US-09-875-195-314

28	20.6	39.6	2641	5	US-09-758-471-1302	Sequence 1302, Ap
c 29	20.6	39.6	5271	6	US-60-278-258-4682	Sequence 4682, Ap
c 30	20.6	39.6	6635	6	US-60-278-232-4263	Sequence 4263, Ap
31	20.6	39.6	7010	5	US-09-764-891-6031	Sequence 6031, Ap
c 32	20.6	39.6	19616	5	US-09-764-877-3220	Sequence 3220, Ap
33	20.6	39.6	93672	5	US-09-803-736-1459	Sequence 1459, Ap
34	20.6	39.6	107931	5	US-09-803-736-752	Sequence 752, App
35	20.4	39.2	288	5	US-09-803-110-6088	Sequence 6088, Ap
36	20.4	39.2	397803	5	US-09-803-110-217	Sequence 217, App
c 37	20.2	38.8	482	5	US-09-823-301-2859	Sequence 2859, Ap
38	20.2	38.8	517	5	US-09-758-460-28	Sequence 28, Appl
c 39	20.2	38.8	254289	5	US-09-803-110-212	Sequence 212, App
c 40	20	38.5	108582	5	US-09-803-736-1236	Sequence 1236, Ap
41	20	38.5	110514	5	US-09-803-736-132	Sequence 132, App
c 42	20	38.5	113172	5	US-09-803-736-113	Sequence 113, App
c 43	20	38.5	118737	5	US-09-803-736-553	Sequence 553, App
44	19.8	38.1	605	6	US-60-255-619-12304	Sequence 12304, A
45	19.8	38.1	755	5	US-09-758-472-1387	Sequence 1387, Ap

ALIGNMENTS

RESULT 1
US-09-804-730-10706
; Sequence 10706, Application US/09804730
; GENERAL INFORMATION:
; APPLICANT: Andersen, Scott E.
; APPLICANT: Hammond-Kosack, Kim
; APPLICANT: Masucci, James D.
; APPLICANT: Urban, Martin
; TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
; FILE REFERENCE: 38-21(51838)B
; CURRENT APPLICATION NUMBER: US/09/804, 730
; CURRENT FILING DATE: 2001-03-13
; PRIOR APPLICATION NUMBER: US 60/189,657
; PRIOR FILING DATE: 2000-03-15
; NUMBER OF SEQ ID NOS: 22828
; SEQ ID NO 10706
; LENGTH: 422
; TYPE: DNA
; ORGANISM: Triticum aestivum
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(422)
; OTHER INFORMATION: unsure at all n locations
; OTHER INFORMATION: Clone ID: LIB3352-021-PI-K1-G5
US-09-804-730-10706

Query Match 45.0%; Score 23.4; DB 5; Length 422;
Best Local Similarity 67.3%; Pred. No. 3.9;
Matches 33; Conservative 0; Mismatches 16; Indels 0; Gaps 0;
QY 3 gcttcgctcactgcggagcgctaaagcgccgccccttaacaaacggtt 51
Db 74 gcttccttcactgcgggcccgcctcccaagcagcgtcctcttccacggaattct 122

RESULT 2
US-09-904-703-1102/c
; Sequence 1102, Application US/09904703
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-758CON1
; CURRENT APPLICATION NUMBER: US/09/904, 703
; CURRENT FILING DATE: 2001-07-12
; PRIOR APPLICATION NUMBER: 09/210, 298
; PRIOR FILING DATE: 1998-12-09
; NUMBER OF SEQ ID NOS: 17812

```
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1102
; LENGTH: 388
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-904-703-1102

Query Match          43.1%; Score 22.4; DB 5; Length 388;
Best Local Similarity 66.7%; Pred. No. 9.5;
Matches 32; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

Qy  2  cgcctgcctcactgcgggaccggtctaaagccggccccccttaacccaacg 49
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 137 CCCTTCACACACTGATGGACACGCTACACAGGAGCGATACAAAGG 90

RESULT 3
US-09-904-703-14754/C
; Sequence 14754, Application US/09904703
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-758CON1
; CURRENT APPLICATION NUMBER: US/09/904,703
; CURRENT FILING DATE: 2001-07-12
; PRIOR APPLICATION NUMBER: 09/210,298
; PRIOR FILING DATE: 1998-12-09
; NUMBER OF SEQ ID NOS: 17812
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 14754
; LENGTH: 531
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)-(531)
; OTHER INFORMATION: n = A,T,C or G
US-09-904-703-14754

Query Match          43.1%; Score 22.4; DB 5; Length 531;
Best Local Similarity 66.7%; Pred. No. 10;
Matches 32; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

Qy  2  cgcctgcctcactgcgggaccggtctaaagccggccccccttaacccaacg 49
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 216 CCCTTCACACACTGATGGACACGCTACACAGGAGCGATACAAAGG 169

RESULT 4
US-09-758-474-201/C
; Sequence 201, Application US/09758474
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PM012
; CURRENT APPLICATION NUMBER: US/09/758,474
; CURRENT FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: 60/179,065
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/180,628
; PRIOR FILING DATE: 2000-02-04
; NUMBER OF SEQ ID NOS: 1402
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 201
; LENGTH: 1409
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (34)

; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (46)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (1382)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (1387)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-758-474-201

Query Match          43.1%; Score 22.4; DB 5; Length 1409;
Best Local Similarity 66.7%; Pred. No. 12;
Matches 32; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

Qy  2  cgcctgcctcactgcgggaccggtctaaagccggccccccttaacccaacg 49
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1072 CCCTTCACACACTGATGGACACGCTACACAGGAGCGATACAAAGG 1025

RESULT 5
US-09-803-110-6363
; Sequence 6363, Application US/09803110
; GENERAL INFORMATION:
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; TITLE OF INVENTION: Agrobacterium tumefaciens Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15490)D
; CURRENT APPLICATION NUMBER: US/09/803,110
; CURRENT FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 09/739,449
; PRIOR FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: US 09/514,000
; PRIOR FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: US 60/168,139
; PRIOR FILING DATE: 1999-12-01
; NUMBER OF SEQ ID NOS: 13351
; SEQ ID NO 6363
; LENGTH: 837
; TYPE: DNA
; ORGANISM: Agrobacterium tumefaciens
US-09-803-110-6363

Query Match          42.7%; Score 22.2; DB 5; Length 837;
Best Local Similarity 69.8%; Pred. No. 13;
Matches 30; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Qy  2  cgcctgcctcactgcgggaccggtctaaagccggccccccttaacc 44
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 96  cgcctgcctgattgcggcagggtgcgcgctgcctctcttgc 138

RESULT 6
US-09-803-110-4177
; Sequence 4177, Application US/09803110
; GENERAL INFORMATION:
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; TITLE OF INVENTION: Agrobacterium tumefaciens Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15490)D
; CURRENT APPLICATION NUMBER: US/09/803,110
; CURRENT FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 09/739,449
; PRIOR FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: US 09/514,000
; PRIOR FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: US 60/168,139
; PRIOR FILING DATE: 1999-12-01
; NUMBER OF SEQ ID NOS: 13351
; SEQ ID NO 4177
```

```
; LENGTH: 1395
; TYPE: DNA
; ORGANISM: Agrobacterium tumefaciens
US-09-803-110-4177

Query Match          42.7%; Score 22.2; DB 5; Length 1395;
Best Local Similarity 69.8%; Pred. No. 14;
Matches 30; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 2 cgttcgtcactcgggaccggtctaaagccggcccttaacc 44
||||| ||| | ||||| | | ||| | |||
Db 410 cgttcgtcgtcgccgaccgggtgacggcggaactgatcc 452

RESULT 7
US-09-764-902-493
; Sequence 493, Application US/09764902
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT213
; CURRENT APPLICATION NUMBER: US/09764,902
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 2318
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 493
; LENGTH: 3157
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (1)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (18)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (20)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (40)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (2469)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (3138)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (3157)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-764-902-493

Query Match          42.7%; Score 22.2; DB 5; Length 3157;
Best Local Similarity 69.8%; Pred. No. 16;
Matches 30; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 3 gcttcgtcactcgggaccggtctaaagccggcccttaacca 45
||||| ||| | ||||| | | ||| | |||
Db 1610 gcttcctcctcgtcgccggtctaaacccggcgccgacgagca 1652

RESULT 8
US-09-764-902-8896/c
; Sequence 8896, Application US/60278232
; GENERAL INFORMATION:
; APPLICANT: Morris, MacDonald
; APPLICANT: Lal, Preeti
; APPLICANT: Diep, Dinh
; TITLE OF INVENTION: Method for the Identification of Sequence Polymorphisms Using

; TITLE OF INVENTION: Polynucleotide Sequence Databases, and Single Nucleotide
; FILE REFERENCE: GX-0011 P
; CURRENT APPLICATION NUMBER: US/60/278,232
; CURRENT FILING DATE: 2001-03-30
; NUMBER OF SEQ ID NOS: 12,557
; SOFTWARE: PERL Program
; SEQ ID NO 8896
; LENGTH: 8416
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; OTHER INFORMATION: Incyte ID No: 402008.3
; NAME/KEY: unsure
; LOCATION: 367-594
; OTHER INFORMATION: a, t, c, g, or other
US-60-278-232-8896

Query Match          42.7%; Score 22.2; DB 6; Length 8416;
Best Local Similarity 69.8%; Pred. No. 18;
Matches 30; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 3 gcttcgtcactcgggaccggtctaaagccggcccttaacca 45
||||| ||| | ||||| | | ||| | |||
Db 4325 GCTTCCTCCTCGTGGCGCGCTAAACCCGGCGCGCAGCAGCA 4283

RESULT 9
US-09-764-902-2125/c
; Sequence 2125, Application US/09764902
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT213
; CURRENT APPLICATION NUMBER: US/09764,902
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 2318
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2125
; LENGTH: 12051
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-902-2125

Query Match          42.7%; Score 22.2; DB 5; Length 12051;
Best Local Similarity 69.8%; Pred. No. 19;
Matches 30; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 3 gcttcgtcactcgggaccggtctaaagccggcccttaacca 45
||||| ||| | ||||| | | ||| | |||
Db 467 GCTTCCTCCTCGTGGCGCGCTAAACCCGGCGCGCAGCAGCA 425

RESULT 10
US-09-764-902-2124/c
; Sequence 2124, Application US/09764902
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT213
; CURRENT APPLICATION NUMBER: US/09764,902
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 2318
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2124
; LENGTH: 18416
; TYPE: DNA
; ORGANISM: Homo sapiens
```

US-09-764-902-2124

Qy 3 gcttcgctcaactgcggagccggctaaagcggcgcccttaacca 45
||||| ||| ||| |||||||| ||||| | |||
Db 467 GCTTCCTTCCTCGCTGGCGCGGCTAAACCGCGCGCAGCAGCA 421

RESULT 11

Query Match 42.7%; Score 22.2; DB 5; Length 260744;

RESULT 12

Query Match

Best Local Similarity 75.0%; Pred. NO. 23;
Matches 27; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 10 tcactgaggacggcgtacaaagcgcccttaacca 45
|| ||||| || ||||| ||||| |||||
Db 1048 tcgctgagcgctacgcaaaagctggccctgaacca 1083

RESULT 15
US-60-278-232-2701
; Sequence 2701, Application US/60278232
; GENERAL INFORMATION:
; APPLICANT: Morris, MacDonald
; APPLICANT: Lal, Preeti
; APPLICANT: Diep, Dinh
; TITLE OF INVENTION: Method for the Identification of Sequence Polymorphisms Using
; TITLE OF INVENTION: Polynucleotide Sequence Databases, and Single Nucleotide
; TITLE OF INVENTION: Polymorphisms Identified Thereby
; FILE REFERENCE: GX-0011 P
; CURRENT APPLICATION NUMBER: US/60/278,232
; CURRENT FILING DATE: 2001-03-30
; NUMBER OF SEQ ID NOS: 12,557
; SOFTWARE: PERL Program
; SEQ ID NO 2701
; LENGTH: 4202
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; OTHER INFORMATION: Incyte ID No: 198653.1
; NAME/KEY: unsure
; LOCATION: 2958-2987
; OTHER INFORMATION: a, t, c, g, or other
US-60-278-232-2701

Query Match 41.5%; Score 21.6; DB 6; Length 4202;
Best Local Similarity 75.0%; Pred. NO. 28;
Matches 27; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 10 tcactgaggacggcgtacaaagcgcccttaacca 45
|| ||||| || ||||| ||||| |||||
Db 3372 tcgctgagcgctacgcaaaagctggccctgaacca 3407

Search completed: August 2, 2001, 16:55:08
Job time: 117 sec

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*

190: gb_est121:*
191: gb_est122:*
192: gb_est123:*
193: gb_est124:*
194: gb_est125:*
195: gb_est126:*
196: gb_est127:*
197: gb_est128:*
198: gb_est129:*
199: gb_est130:*
200: gb_est131:*
201: gb_est132:*
202: gb_est133:*
203: gb_est134:*
204: gb_est135:*
205: gb_est136:*
206: gb_est137:*
207: gb_est138:*
208: gb_est139:*
209: gb_est140:*
210: gb_est141:*
211: gb_est142:*
212: gb_est143:*
213: gb_est144:*
214: gb_est145:*
215: gb_est146:*
216: gb_est147:*
217: gb_est148:*
218: gb_est149:*
219: gb_est150:*
220: gb_est151:*
221: gb_est152:*
222: gb_est153:*
223: gb_est154:*
224: gb_est155:*
225: gb_est156:*
226: gb_est157:*
227: gb_est158:*
228: gb_est159:*
229: gb_est160:*
230: gb_est161:*
231: gb_est162:*
232: gb_est163:*
233: gb_est164:*
234: gb_est165:*
235: gb_est166:*
236: gb_est167:*
237: gb_est168:*
238: gb_est169:*
239: gb_est170:*
240: gb_est171:*
241: gb_est172:*
242: gb_est173:*
243: gb_est174:*
244: gb_est175:*
245: gb_est176:*
246: gb_est177:*
247: gb_est178:*
248: gb_est179:*
249: gb_est180:*
250: gb_est181:*
251: gb_est182:*
252: gb_est183:*
253: gb_est184:*
254: gb_est185:*
255: gb_est186:*
256: gb_est187:*
257: gb_est188:*
258: gb_est189:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query No.	Score	Query		Length	DB	ID	Description
			Match	%				
C	1	23.8	45.8	553	21	AI520326	AI520326	LD40453.5
	2	23.8	45.8	563	21	AI517070	AI517070	GH27702.5
	3	23.8	45.8	991	219	CNS01320	CNS01320	Drosophil
	4	23.6	45.4	1004	220	CNS027PB	CNS027PB	AI184952 Tetraodon
	5	23.4	45.0	2002	1	AA013941	AA013941	mh05c03.1
	6	23.4	45.0	593	141	BE853438	BE853438	uw31f08.x
	7	23.4	45.0	879	221	CNS03G7Z	CNS03G7Z	AL242648 Tetraodon
	8	23	44.2	197	246	A2578481	A2578481	23n08 Shon
	9	22.8	43.8	342	163	BE074183	BE074183	AZ578481 23n08 Shon
	10	22.8	43.8	637	139	BE735159	BE735159	RC1-BT056
C	11	22.8	43.8	815	147	BG116175	BG116175	BE735159 60156901.x
	12	22.6	43.5	402	147	BF374103	BF374103	602318362
	13	22.6	43.5	509	168	BF726683	BF726683	602318362
	14	22.6	43.5	515	228	A0435257	A0435257	py10408.y
	15	22.6	43.5	924	153	EG390246	EG390246	HS-5140.B
	16	22.4	43.1	347	116	AW439871	AW439871	602416203
	17	22.4	43.1	386	19	AI368921	AI368921	qt3f06.x
	18	22.4	43.1	386	22	A1632338	A1632338	qv98c07.x
	19	22.4	43.1	395	20	A1434153	A1434153	tt22904.x
	20	22.4	43.1	406	23	A0653217	A0653217	tt22904.x
C	21	22.4	43.1	492	16	A1126769	A1126769	wb23h11.x
	22	22.4	43.1	518	116	AW440490	AW440490	qb85h08.x
	23	22.4	43.1	546	190	W19597	W19597	qb85h08.x
	24	22.4	43.1	574	10	A644914	A644914	hb91c05.x
	25	22.4	43.1	584	144	BF107283	BF107283	zb36b12.r1
	26	22.4	43.1	635	145	BF167553	BF167553	vs84d04.r
	27	22.4	43.1	643	249	AZ807425	AZ807425	601824180
	28	22.4	43.1	716	10	A3671918	A3671918	601774261
	29	22.4	43.1	731	105	AL110412	AL110412	601774261
	30	22.2	42.7	234	147	BF365198	BF365198	v111002.r
C	31	22.2	42.7	373	104	AZ80017	AZ80017	DKFP434K
	32	22.2	42.7	433	225	AQ174469	AQ174469	PM1-NT001
	33	22.2	42.7	616	31	AV540656	AV540656	4A3A-AAK-
	34	22.2	42.7	670	256	B17432	B17432	HS-3215.A
	35	22	42.3	519	222	FR0013145	FR0013145	AV540656
	36	22	42.3	541	20	A1463539	A1463539	345U24.TVB
	37	22	42.3	866	239	AZ207121	AZ207121	F.rubrripe
	38	22	42.3	991	256	B11618	B11618	mu75e01.x
	39	21.8	41.9	310	108	AV019157	AV019157	SP-0127.SP
	40	21.8	41.9	343	11	AA763260	AA763260	SP-0127.SP
C	41	21.8	41.9	412	146	BF226128	BF226128	T8013-Sp6.3
	42	21.8	41.9	419	115	AW412398	AW412398	u74h06.Y
	43	21.8	41.9	435	165	BE290435	BE290435	60108886
	44	21.8	41.9	458	4	AA266758	AA266758	u74h06.Y
	45	21.8	41.9	478	166	BE303443	BE303443	mz93h06.r

ALIGNMENTS

RESULT	1	
LOCUS	AI520326/c	
DEFINITION	LD04053.5prime LD Drosophila melanogaster embryo pOT2 Drosophila melanogaster cDNA clone LD04053 5prime, mRNA sequence.	
ACCESSION	AI520326	
VERSION	AI520326.1	
KEYWORDS	EST, GI:4426180	
SOURCE	fruit fly, Drosophila melanogaster	
ORGANISM	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.	
REFERENCE	1 (bases 1 to 553)	
AUTHORS	Harvey,D., Hong,L., Evans,Holm,M., Pendleton,J., Su,C., Brokstein ,P., Lewis,S. and Rubin,G.M.	
TITLE	BDGP/HHMI Drosophila EST project	

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Fax: 314 286 1810
Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LLNL ; contact the


```

Query Match      45.0%; Score 23.4; DB 221; Length 879;
Best Local Similarity 73.2%; Pred. No. 1.1e+02;
Matches 30; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 5 ttgcgtcactgcggaccggtctaaagccggcccttaacca 45
    ||||| | ||||| ||||| || | |||||
Db 232 TTGCACCTTTGGCGGACCGGCTAAACACCACCGATTGAACCA 192

RESULT 8
LOCUS A2578481/c 197 bp DNA GSS 08-DEC-2000
DEFINITION 23h08 Shot-gun genomic library of Rhizobium strain ANU265 Rhizobium
sp. NGR234 genomic clone 23h08, DNA sequence.
ACCESSION A2578481
VERSION A2578481.1 GI:11605806
KEYWORDS GSS.
SOURCE Rhizobium sp. NGR234.
ORGANISM Rhizobium sp. NGR234
Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
Rhizobiaceae; Rhizobium.
REFERENCE 1 (bases 1 to 197)
AUTHORS Viprey,V., Rosenthal,A., Broughton,W.J. and Perret,X.
TITLE Genetic snapshots of the Rhizobium species NGR234 genome
JOURNAL Genome Biology, vol 1 (6), 0014.1-0014.7 (2000)
COMMENT Contact: Virginie Viprey
Laboratoire de Biologie Moleculaire des Plantes Superieures
University of Geneva
1 Chemin de l'Imperatrice, Chambesy/Geneva 1292, Switzerland
Tel: +44(0)1603450000
Fax: +44(0)1603450045
Email: virginie.viprey@bbrc.ac.uk
Class: shotgun.
FEATURES
    Location/Qualifiers
    1..197
    /organism="Rhizobium sp. NGR234"
    /strain="ANU265"
    /db_xref="taxon:394"
    /clone="23h08"
    /clone_lib="Shot-gun genomic library of Rhizobium strain
    ANU265."
    /note="Vector: M13; derivative strain of NGR234 cured of
    pNGR234a"
BASE COUNT 37 a 73 c 62 g 24 t 1 others
ORIGIN

Query Match      44.2%; Score 23; DB 246; Length 197;
Best Local Similarity 74.4%; Pred. No. 1.4e+02;
Matches 29; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 1 tcgcttcactgcggaccggtctaaagccggccct 39
    ||||| ||||| ||||| ||||| ||||| |||||
Db 166 TTGCTTCTCAAGCGGGTTCGGCGACGCGCCGTCGCT 128

RESULT 9
LOCUS BE074183 342 bp mRNA EST 09-JUN-2000
DEFINITION RCI-BT0567-301299-011-c04 BT0567 Homo sapiens cDNA, mRNA sequence.
ACCESSION BE074183
VERSION BE074183.1 GI:8421866
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 342)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and


```

```

Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?cl=62-RCL-BT0567-301
299-011-c04&t3=1999-12-30&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 342.
Location/Qualifiers
    1..342
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone_lib="BT0567"
    /dev_stage="Adult"
    /note="Organ: breast; Vector: puc18; Site_1: Sma1; Site_2:
    Sma1; A mini-library was made by cloning products derived
    from ORESTES PCR (U.S. Letters Patent application No. 196
    ,716 - Ludwig Institute for Cancer Research) profiles
    into the puc 18 vector. Reverse transcription of tissue
    mRNA and cDNA amplification were performed under low
    stringency conditions."
BASE COUNT 77 a 89 c 87 g 89 t
ORIGIN

```

```

Query Match      43.8%; Score 22.8; DB 163; Length 342;
Best Local Similarity 71.4%; Pred. No. 1.7e+02;
Matches 30; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 11 cactgcgggaccgctaaagccggcccttaacaaacgtta 52
    ||||| | | |||| | |||| | |||| | |||| | |||| |
Db 230 CCCTGCGGTAGTGCATATAAACCTGCACCAACAAACGTTA 271

```

```

RESULT 10
LOCUS BE735159 637 bp mRNA EST 15-SEP-2000
DEFINITION 601569107F1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:3844009 5',
mRNA sequence.
ACCESSION BE735159
VERSION BE735159.1 GI:10149151
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 637)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cyapbs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LLCM539 row: m column: 02
High quality sequence start: 3
High quality sequence stop: 589.
Location/Qualifiers

```

```

source
1. .637
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3844009"
/clone_lib="NIH_MGC_21"
/tissue_type="choriocarcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: placenta; Vector: pOTB7; Site_1: XhoI;
Site_2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA Synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
```

BASE COUNT 159 a 199 c 157 g 122 t

ORIGIN

```

Query Match 43.8%; Score 22.8; DB 139; Length 637;
Best Local Similarity 71.4%; Pred. No. 1.7e+02;
Matches 30; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 6 tcgctcactgcggaccgctaaagccgccccttaacacaa 47
||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 492 TCAGTCACTACTGGAAGTCTCAAACTGGCCACTGAACAAA 533

RESULT 11
BG116175 815 bp mRNA EST 30-JAN-2001
LOCUS 602318362F1 NIH_MGC_88 Homo sapiens cDNA clone IMAGE:4418822 5',
mRNA sequence.
ACCESSION BG116175
VERSION BG116175.1 GI:12609681
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 815)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs@email.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10153 row: k column: 15
High quality sequence stop: 733.
FEATURES
source
Location/Qualifiers
1. .815
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4418822"
/clone_lib="NIH_MGC_88"
/tissue_type="duodenal adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: small intestine; Vector: pCMV-SPORT6;
Site_1: NotI; Site_2: SalI; Cloned unidirectionally;
oligo-dT primed. Average insert size 1.767 kb. Library
enriched for full-length clones and constructed by Life
Technologies. Note: this is a NIH_MGC Library."
```

BASE COUNT 189 a 267 c 202 g 157 t

ORIGIN

```

Query Match 43.8%; Score 22.8; DB 174; Length 815;
Best Local Similarity 71.4%; Pred. No. 1.7e+02;
Matches 30; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 6 tcgctcactgcggaccgctaaagccgccccttaacacaa 47
||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 492 TCAGTCACTACTGGAAGTCTCAAACTGGCCACTGAACAAA 533

RESULT 11
BG116175 815 bp mRNA EST 30-JAN-2001
LOCUS 602318362F1 NIH_MGC_88 Homo sapiens cDNA clone IMAGE:4418822 5',
mRNA sequence.
ACCESSION BG116175
VERSION BG116175.1 GI:12609681
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 815)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs@email.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10153 row: k column: 15
High quality sequence stop: 733.
FEATURES
source
Location/Qualifiers
1. .815
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4418822"
/clone_lib="NIH_MGC_88"
/tissue_type="duodenal adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: small intestine; Vector: pCMV-SPORT6;
Site_1: NotI; Site_2: SalI; Cloned unidirectionally;
oligo-dT primed. Average insert size 1.767 kb. Library
enriched for full-length clones and constructed by Life
Technologies. Note: this is a NIH_MGC Library."
```

BASE COUNT 189 a 267 c 202 g 157 t

ORIGIN

```

Best Local Similarity 71.4%; Pred. No. 1.8e+02;
Matches 30; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 6 tcgctcactgcggaccgctaaagccgccccttaacacaa 47
||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 544 TCATCACTACTGGAAGTCTCAAACTGGCCACTGAACAAA 585

RESULT 12
BF374103 402 bp mRNA EST 24-NOV-2000
LOCUS MR3-SN0066-020500-003-g01_1 SN0066 Homo sapiens cDNA, mRNA
DEFINITION sequence.
ACCESSION BF374103
VERSION BF374103.1 GI:11336232
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 402)
AUTHORS Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M. R.,
Nagai, M. A., da Silva, W. Jr., Zago, M. A., Bordin, S., Costa, F. F.,
Goldman, G. H., Carvalho, A. F., Matsukuma, A., Baia, G. S., Simpson, D. H.,
Brunstein, A., de Oliveira, P. S., Bucher, P., Jongeneel, C. V., O'Hare,
M. J., Soares, F., Brentani, R. R., Reis, L. F., de Souza, S. J., and
Simpson, A. J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=MR3&tl2=MR3-SN0066-
020500-003-g01.1&tl3=2000-05-02&tl4=1)
Seq primer: puc 18 forward
FEATURES
source
Location/Qualifiers
1. .402
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="SN0066"
/dev_stage="Adult"
/note="Organ: stomach_normal; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
```

BASE COUNT 92 a 103 c 86 g 121 t

ORIGIN

```

Query Match 43.5%; Score 22.6; DB 147; Length 402;
Best Local Similarity 68.9%; Pred. No. 2e+02;
Matches 31; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 8 gctcactgcggaccgctaaagccgccccttaacacacgta 52
||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2 GCTCATTCAGGACCGCTCATGTCCAGCGCTTACCTTACCTTA 46

RESULT 13
BF726683/c 509 bp mRNA EST 05-JAN-2001
LOCUS BF726683
```

```

DEFINITION   by10d08.y1 Human Lens cDNA (Un-normalized, unamplified): BY Homo
ACCESSION    BF726683
VERSION      BF726683.1 GI:12042594
KEYWORDS     EST.
SOURCE       human.
ORGANISM     Homo sapiens
REFERENCE    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE        Wistow,G.J., Bernstein,S., Behal,A. and Smith,D.
JOURNAL      NEIBANK: EST analysis and bioinformatics for ocular genomics
COMMENT      Invest. Ophthalmol. Vis. Sci. 41 (2000) In press
              Contact: Wistow G
              Section on Molecular Structure and Function
              National Eye Institute
              6/331, NIH, Bethesda, MD 20892-2740, USA
              Tel: 301 402 3452
              Fax: 301 496 0078
              Email: graeme@helix.nih.gov
              Plate: 10 row: d column: 08
              Seq primer: M13Rpl reverse primer (ABI).
FEATURES     Location/Qualifiers
              1..509
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              /clone="by10d08"
              /clone_lib="Human Lens cDNA (Un-normalized, unamplified):
              BY"
              /tissue_type="Lens"
              /dev_stage="Adult"
              /lab_host="EMDH108"
              /note="Organ: Eye; Vector: pCMVSPORT6; Two human lenses
              from different adults (both approximately 40 years old)
              together yielded 20ug of total RNA and 150ng mRNA for cDNA
              library synthesis. A directionally cloned cDNA library in
              the pCMVSPORT6 vector was constructed at Life Technologies
              , essentially following the protocols of the SuperScript
              Plasmid System full details of which are contained in the
              manufacturer's instruction manual
              (http://www.lifetech.com/). First strand synthesis was
              carried out using a Not I primer-adaptor
              [5'-pGACTAGTTCTAGATCGGAGCGGCCCTT15-3']. Not I/blunt
              end inserts were cloned into the Not I/EcoR V sites in the
              vector. EST analysis was performed on the unamplified
              library at the NIH Intramural Sequencing Center (NISC)."
BASE COUNT   95 a 144 c 156 g 113 t 1 others
ORIGIN
Query Match  43.5%; Score 22.6; DB 168; Length 509;
Best Local Similarity 68.9%; Pred. No. 2e+02;
Matches 31; Conservative 0; Mismatches 14; Indels 0; Gaps 0;
QY 4 cttcgtcactcgggaccgctaaagccggcccttaaccaaac 48
    ||||| 1 ||||| ||||| ||||| ||||| |||||
Db 95 CTCTCAGACTCGGGGACCTGCAGCACCCTCGCCCGCCAGTCTCTGAAC 51

RESULT 14
LOCUS       A0435257
DEFINITION HS_5140_B2_H04_SP6E RPCI-11 Human Male BAC Library Homo sapiens
            genomic clone Plate=716 Col=8 Row=P, DNA sequence.
ACCESSION  A0435257
VERSION    A0435257.1 GI:4546596
KEYWORDS   GSS.
SOURCE     human.
ORGANISM   Homo sapiens
REFERENCE  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
            Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,

us-09-701-626a-79.rst
DEFINITION   Sequence-tagged connectors: A sequence approach to mapping and
              scanning the human genome
              Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
              95380589
              Contact: Mahairas GG, Wallace JC, Hood L
              High Throughput Sequencing Center
              University of Washington
              401 Queen Anne Avenue North, Seattle, WA 98109, USA
              Tel: (206) 616-3618
              Fax: (206) 616-3887
              Email: jwallace@u.washington.edu
              Clones are derived from the human BAC library RPCI-11. For BAC
              library availability, please contact Pieter de Jong
              (pieter@dejong.med.buffalo.edu). Clones may be purchased from
              BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
              or from Resear h Genetics (info@resgen.com). BAC end Web Server:
              http://www.htsc.washington.edu
              Plate: 716 row: P column: 8
              Seq primer: SP6
              Class: BAC ends
              High quality sequence stop: 515.
FEATURES     Location/Qualifiers
              1..515
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              /clone="Plate=716 Col=8 Row=P"
              /clone_lib="RPCI-11 Human Male BAC Library"
              /sex="male"
              /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
              Male blood DNA was isolated from one randomly chosen donor
              and partially digested with a combination of EcoRI and
              EcoRI Methylase. Size selected DNA was cloned into the
              pBACe3.6 vector at EcoRI sites"
BASE COUNT   154 a 88 c 91 g 179 t 3 others
ORIGIN
Query Match  43.5%; Score 22.6; DB 228; Length 515;
Best Local Similarity 68.9%; Pred. No. 2e+02;
Matches 31; Conservative 0; Mismatches 14; Indels 0; Gaps 0;
QY 8 gctcactcgggaccgctaaagccggcccttaaccaaacgtta 52
    ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 263 GATTTCTGCTGACAGCTTTAAAGCAGCCCTCCTAACACAAAAGATA 307

RESULT 15
LOCUS       BG390246
DEFINITION 602416203F1 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:4524648 5',
            mRNA sequence.
ACCESSION  BG390246
VERSION    BG390246.1 GI:13283694
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
REFERENCE  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
            NIH-MGC http://mgc.nci.nih.gov/.
            National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished (1999)
            Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-remail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov

```


Plate: LLAM10429 row: e column: 01
High quality sequence stop: 718.
Location/Qualifiers
1. .924
source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4524648"
/clone_lib="NIH_MGC_92"
/tissue_type="embryonal carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: testis; Vector: pCMV-SPORT6; Site.1: NotI;
Site.2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 2.5 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC library."
BASE COUNT 265 a 186 c 162 g 311 t
ORIGIN

Query Match 43.5%; Score 22.6; DB 153; Length 924;
Best Local Similarity 75.7%; Pred. No. 2.le+02;
Matches 28; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
Qy 16 cgggaccggttaagcgccgcttaacccaacgtta 52
||||||| ||| ||| ||| ||| ||| ||| |||
Db 814 CGGACCCGCTTAACCGGTCCTTAACCCAAACGTAA 850

Search completed: August 2, 2001, 18:39:11
Job time: 6361 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 2, 2001, 18:06:17 ; Search time 1350.43 Seconds
(without alignments)
595.605 Million cell updates/sec

Title: US-09-701-626A-79
Perfect score: 52
Sequence: 1 tcgtcttcgtcactgcggga.....ggcccttaaccaaactgta 52

Scoring table: IDENTITY_NUC
Gapop 10_0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues

Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0
Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*

- 1: gb_ba1:*
- 2: gb_ba2:*
- 3: gb_ba3:*
- 4: gb_in1:*
- 5: gb_in2:*
- 6: gb_in3:*
- 7: gb_om:*
- 8: gb_ov:*
- 9: gb_pat1:*
- 10: gb_pat2:*
- 11: gb_ph:*
- 12: gb_p11:*
- 13: gb_p12:*
- 14: gb_p13:*
- 15: gb_p14:*
- 16: em_ba1:*
- 17: em_ba2:*
- 18: em_fun:*
- 19: em_htgo_hum:*
- 20: em_htgo_inv:*
- 21: em_htgo_rod:*
- 22: em_htg_hum1:*
- 23: em_htg_hum2:*
- 24: em_htg_hum3:*
- 25: em_htg_hum4:*
- 26: em_htg_hum5:*
- 27: em_htg_hum6:*
- 28: em_htg_hum7:*
- 29: em_htg_hum8:*
- 30: em_htg_inv1:*
- 31: em_htg_inv2:*
- 32: em_htg_other:*
- 33: em_htg_rod:*
- 34: em_hum1:*
- 35: em_hum2:*
- 36: em_hum3:*
- 37: em_hum4:*
- 38: em_hum5:*
- 39: em_hum6:*
- 40: em_hum7:*
- 41: em_in:*
- 42: em_om:*
- 43: em_or:*

44: em_ov:*

45: em_pat:*

46: em_ph:*

47: em_pl:*

48: em_ro:*

49: em_sts:*

50: em_sy:*

51: em_un:*

52: em_vi:*

53: gb_sts1:*

54: gb_sts2:*

55: gb_sts3:*

56: gb_sy:*

57: gb_un:*

58: gb_vil:*

59: gb_vil2:*

60: gb_htg1:*

61: gb_htg2:*

62: gb_htg3:*

63: gb_htg4:*

64: gb_htg5:*

65: gb_htg6:*

66: gb_htg7:*

67: gb_htg8:*

68: gb_htg9:*

69: gb_htg10:*

70: gb_htg11:*

71: gb_htg12:*

72: gb_htg13:*

73: gb_htg14:*

74: gb_htg15:*

75: gb_htg16:*

76: gb_htg17:*

77: gb_htg18:*

78: gb_htg19:*

79: gb_htg20:*

80: gb_htg21:*

81: gb_htg22:*

82: gb_htg23:*

83: gb_htg24:*

84: gb_htg25:*

85: gb_pr1:*

86: gb_pr2:*

87: gb_pr3:*

88: gb_pr4:*

89: gb_pr5:*

90: gb_pr6:*

91: gb_pr7:*

92: gb_pr8:*

93: gb_pr9:*

94: gb_rol:*

95: gb_rol2:*

96: gb_in4:*

97: gb_pr10:*

98: em_ba3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
C 1	24.2	46.5	1712	3 PSELASA	M20982 P.aeruginos
C 2	24.2	46.5	3742	3 PAU68175	U68175 Pseudomonas
C 3	23.8	45.8	4943	63 AC014389	AC014389 Drosophil
C 4	23.8	45.8	294218	4 AE003430	AE003430 Drosophil
5	23.6	45.4	9581	3 PAE223604	AJ223604 Pseudomon
6	23.2	44.6	68596	69 AC025118	AC025118 Homo sapi
7	23.2	44.6	181052	91 CNS01DWP	AL137818 Human chr
8	23	44.2	33615	2 BSU11039	U11039 Bacillus su

```
9 23 44.2 120766 85 AC004150
10 23 44.2 208780 2 BSUB0009
11 23 44.2 218864 68 AC023855
12 23 44.2 233780 2 BSUB0010
13 22.8 43.8 172569 79 AL353608
14 22.6 43.5 1142 12 AF033487
15 22.6 43.5 45257 76 AC084177
16 22.6 43.5 95255 88 AF051976
17 22.6 43.5 133075 78 AF205589
18 22.6 43.5 163608 69 AC026081
19 22.6 43.5 173749 66 AC020567
20 22.4 43.1 1233 94 MNCERBA1
21 22.4 43.1 1240 94 MNCERBA3
22 22.4 43.1 1357 94 MNCERBA2
23 22.4 43.1 1440 94 MNCERBA
24 22.4 43.1 167294 63 AC013814
25 22.4 43.1 179023 73 AC068768
26 22.2 42.7 173 9 AR095503
27 22.2 42.7 303 58 AB026094
28 22.2 42.7 303 58 AB026096
29 22.2 42.7 303 58 AB026097
30 22.2 42.7 303 58 AB026098
31 22.2 42.7 775 58 AF075212
32 22.2 42.7 822 58 AF058736
33 22.2 42.7 117620 82 AL590618
34 22.2 42.7 128396 71 AC034262
35 22.2 42.7 132805 92 HS339A18
36 22.2 42.7 137604 67 AC022275
37 22.2 42.7 148204 74 AC069400
38 22.2 42.7 164630 66 AC020632
39 22.2 42.7 171744 83 CNS01DW9
40 22.2 42.7 180238 74 AC073688
41 22.2 42.7 182741 65 AC018910
42 22.2 42.7 184825 90 AL391380
43 22.2 42.7 195705 68 AC024270
44 22.2 42.7 224592 75 AC074209
45 22 42.3 591 54 ECU90592
```

ALIGNMENTS

```
RESULT 1
PSLASA/c 1712 bp DNA BCT 26-APR-1993
LOCUS P.aeruginosa structural elastase protein (lasA) gene, complete cds.
DEFINITION M20982.1 GI:151323
KEYWORDS lasA gene; structural elastase protein.
SOURCE P.aeruginosa DNA, clone pPS1816.
ORGANISM Pseudomonas aeruginosa
Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
Pseudomonas.
1 (bases 1 to 1712)
Schad,P.A. and Iglewski,B.H.
Nucleotide sequence and expression in Escherichia coli of the
Pseudomonas aeruginosa lasA gene
J. Bacteriol. 170, 2784-2789 (1988)
MEDLINE 8827864
COMMENT This protein will react with elastase antibodies but shows no
elastase activity.
FEATURES
Location/Qualifiers
1..1712
/organism="Pseudomonas aeruginosa"
/db_xref="taxon:287"
148..152
RBS /note="S-D ribosomal binding site (put.); putative"
159..1292
CDS /note="structural elastase"
/codon_start=1
/transl_table=11
/protein_id="AAA25873.1"
/db_xref="GI:151324"
```

```
/translation="MOHKRRAMASPRSPFLFVLALAVGTTANAHDDGLPAFRYSAE
LLGQLQPSVALPLNDLFLYGRDAEAFDEALAYLALNAPALRDKSEYLEHWSGYYSIN
PKVLLTLVMQSGPLGADERLAARGLRSKARGFDAQVRDVLQOLRSRYGFEFYQ
LRQAAARXAVGEDGLNAAASALLGLREGAKYSAVARRQSAAYAQTFORLFGTFAAE
LLOPSNRVAROLOAKAALAPPSNLMOLPWRQGYSHQPNGAHFEGSGYPSYSFDSYD
WPRWGSATYSVVAHAAGTVRVLRSQVRVTHPSGNAATNYHHMDQIQVNSGQVSADTK
LGVIAGNTALTACGGSGTGPLHFLSLYNGAFVSLQ...SFGP"
BASE COUNT 276 a 596 c 561 g 279 t
ORIGIN 17 bp upstream of SmaI site.

Query Match 46.5%; Score 24.2; DB 3; Length 1712;
Best Local Similarity 78.4%; Pred. No. 1.7e+02;
Matches 29; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 2 cgccttcgctcactgcggaccgctaaagccgccc 38
||| ||||| ||| ||||| ||||| ||||| |||
DB 1463 CGTTTCGCTCGCGGCGGCGGCTGGAGCCGCGCCC 1427

RESULT 2
PAU68175/c
LOCUS PAU68175 3742 bp DNA BCT 06-APR-1998
DEFINITION Pseudomonas aeruginosa staphylytic protease preproenzyme LasA
(lasA) gene, complete cds.
ACCESSION U68175
VERSION U68175.1 GI:1549396
KEYWORDS Pseudomonas aeruginosa.
SOURCE Pseudomonas aeruginosa
ORGANISM Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
Pseudomonas.
1 (bases 1 to 3742)
Gustin,J.K., Kessler,E. and Ohman,D.E.
A substitution at His-120 in the LasA protease of Pseudomonas
aeruginosa blocks enzymatic activity without affecting propeptide
processing or extracellular secretion
J. Bacteriol. 178 (22), 6608-6617 (1996)
MEDLINE 97086629
REFERENCE 2 (bases 1 to 3742)
Gustin,J.K., Kessler,E. and Ohman,D.E.
Direct Submission
TITLE Submitted (27-AUG-1996) Microbiology and Immunology, University of
Tennessee and Veterans Affairs Medical Center, 858 Madison Ave.,
Room 101, Memphis, TN 38163, USA
JOURNAL Location/Qualifiers
1..3742
/organism="Pseudomonas aeruginosa"
/strain="FRD1"
/db_xref="taxon:287"
/map="55 min."
658..660
misc_feature /note="transcriptional start site"
evidence-not_experimental
gene 763..2030
/organism="lasA"
RBS 763..767
/organism="lasA"
774..866
sig_peptide /gene="lasA"
774..2030
CDS /gene="lasA"
/codon_start=1
/transl_table=11
/product="staphylytic protease preproenzyme LasA"
/protein_id="AAC12656.1"
/db_xref="GI:1549397"
/translation="MOHKRRAMASPRSPFLFVLALAVGTTANAHDDGLPAFRYSAE
LLGQLQPSVALPLNDLFLYGRDAEAFDEALAYLALNAPALRDKSEYLEHWSGYYSIN
PKVLLTLVMQSGPLGADERLAARGLRSKARGFDAQVRDVLQOLRSRYGFEFYQ
LRQAAARXAVGEDGLNAAASALLGLREGAKYSAVARRQSAAYAQTFORLFGTFAAE
LLOPSNRVAROLOAKAALAPPSNLMOLPWRQGYSHQPNGAHFEGSGYPSYSFDSYD
WPRWGSATYSVVAHAAGTVRVLRSQVRVTHPSGNAATNYHHMDQIQVNSGQVSADTK
```

```
LGVYAGNINTALCEGSSGPHLHFLSYLLNGAFVSILQASFGPYRINVGTSGSYNDNCR
RYFYNQSGATTHCAFRPLYNPGLAL"
/genes="lasa"
/notes="staphylyolytic protease"
1839..1841
/genes="lasa"
/notes="encodes histidine residue critical for enzymatic
activity"
complement(2144..2935)
/notes="ORF1"
/codon_start=1
/transl_table=11
/db_xref="GI:1549398"
/translation="MSLSPADLKQLQAANKLLENPGLTAQLTSLTGTPTEYGLKLPLR
KYNEKIGSVIEGALLKVARSATATLQENASGOPANRLHTLGAASGVGGFGFEPGLAML
VEVPVTTGIIIFRSIADIARSEGESIRDMOTLMACIEVFALGGKASDDASGESYAVR
TALAOOVKAADFLCKDAGSKAAKPLMLALVRKVAERLGVQYSEKLAQIVPIVGALG
GAANTVFNNHFQAMRGHFVVRRLERHYGKDVVRDAYALLPRG"
BASE COUNT 631 a 1236 c 1278 g 597 t
ORIGIN

Query Match 46.5%; Score 24.2; DB 3; Length 3742;
Best Local Similarity 78.4%; Pred. No. 1.4e+02;
Matches 29; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 2 cgcctgcctcactcgccgacgcggtacgaagccgcgcc 38
II IIIIIII I IIII IIIII IIIII IIII IIII
Db 2076 CGTTTCGCTCGCGCGGAGCGCGGTGAGCGGGGCC 2040

RESULT 3
AC014389/c
LOCUS
DEFINITION AC014389 4943 bp DNA HTG 16-NOV-1999
Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
pieces.
AC014389
VERSION AC014389.1 GI:6436946
KEYWORDS HTG; HTGS_PHASE2.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 4943)
REFERENCE Adams, M. and Venter, J.C.
AUTHORS Direct Submission
TITLE Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive,
JOURNAL Rockville, MD, USA
COMMENT This sequence was identified as CDM:10214250 by the submitter.
For further information on this sequence e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES
source
1..4943
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
BASE COUNT 1518 a 1061 c 1096 g 1268 t
ORIGIN

Query Match 45.8%; Score 23.8; DB 63; Length 4943;
Best Local Similarity 72.1%; Pred. No. 1.8e+02;
Matches 31; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Qy 2 cgcctgcctcactcgccgacgcggtacgaagccgcgcccttaacc 44
II IIII IIII IIII IIIII IIII IIII IIII IIII
Db 1092 CGATTTCGGACCAAGGCGAGCGGTAAAGCCGGCCATTAAAC 1050
```

```
RESULT 4
AE003430/c
LOCUS
DEFINITION
Drosophila melanogaster genomic scaffold 142000013386054 section 14
of 35, complete sequence.
ACCESSION AE003430
VERSION AE003430
KEYWORDS AE003430
SOURCE GI:10728458
ORGANISM HTG.
fruit fly.
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 294218)
REFERENCE Adams, M.D., Celniker, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D.,
AUTHORS Ananatiadis, P.G., Scherer, S.E., Li, P.W., Hoskins, R.A., Galie, K.F.,
George, R.A., Lewis, S.E., Richards, S., Ashburner, M., Henderson, S.N.,
Sutton, G.G., Wortman, J.R., Yandell, M.D., Zhang, Q., Chen, L.X.,
Brandon, R.C., Rogers, J.H., Blazej, R.G., Champe, M., Pfeiffer, B.D.,
Wan, K.H., Doyle, C., Baxter, E.G., Helt, G., Nelson, C.R., Gabor
Miklos, G.L., Abail, J.F., Agbayani, A., An, H.J.,
Andrews-Pfannkoch, C., Baldwin, D., Ballew, R.M., Basu, A.,
Baxendale, J., Bayraktaroglu, L., Beasley, E.M., Beeson, K.Y.,
Benos, P.V., Berman, B.P., Bhandari, D., Bolshakov, S., Borkova, D.,
Botchan, M.R., Bouck, J., Broksstein, P., Brottier, P., Burtis, K.C.,
Busam, D.A., Butler, H., Cadieu, E., Center, A., Chandra, I.,
Cherry, J.M., Cawley, S., Dahlke, C., Davenport, L.B., Davies, P., de
Pablos, B., Delcher, A., Deng, Z., Mays, A.D., Dew, I., Dietz, S.M.,
Dodson, K., Doup, L.E., Downes, M., Dugan-Rocha, S., Dunkov, B.C.,
Dunn, P., Durbin, K.J., Evangelista, C.C., Ferraz, C., Ferrier, S.,
Fleischmann, W., Fosler, C., Gabrielian, A.E., Garg, N.S.,
Gelbart, W.M., Glasser, K., Glodek, A., Gong, F., Gorrell, J.H., Gu, Z.,
Guan, P., Harris, M., Harris, N.L., Harvey, D., Heiman, T.J.,
Hernandez, J.R., Houck, J., Hostin, D., Houston, K.A., Howland, T.J.,
Wei, M.H., Ibegwam, C., Jalali, M., Kalush, F., Karpen, G.H., Ke, Z.,
Kennison, J.A., Ketchum, K.A., Kimmel, B.E., Kodira, C.D., Kraft, C.,
Kravitz, S., Kulp, D., Lai, Z., Lasko, P., Lei, Y., Levitsky, A.A.,
Li, J., Li, Z., Liang, Y., Lin, X., Liu, X., Mattel, B., McIntosh, T.C.,
McLeod, M.P., McPherson, D., Merkulov, G., Milshina, N.V., Mobarry, C.,
Morriss, J., Moshrefi, A., Mount, S.M., Moy, M., Murphy, B., Murphy, L.,
Muzny, D.M., Nelson, D.L., Nelson, D.R., Nelson, K.A., Nixon, K.,
Nusskern, D.R., Pacleb, J.M., Palazzolo, M., Pittman, G.S., Pan, S.,
Pollard, J., Puri, V., Reese, M.G., Reinert, K., Remington, K.,
Saunders, R.D., Scheeler, F., Shen, H., Shue, B.C., Siden-Kiamos, I.,
Simpson, M., Skupski, M.P., Smith, T., Spier, E., Spradling, A.C.,
Stapleton, M., Strong, R., Sun, E., Svirska, R., Tector, C., Turner, R.,
Venter, E., Wang, A.H., Wang, X., Wang, Z.Y., Wassarman, D.A.,
Weinstock, G.M., Weissbach, J., Williams, S.M., Woodage, T.,
Worley, K.C., Wu, D., Yang, S., Yao, Q.A., Ye, J., Yeh, R.F.,
Zaveri, J.S., Zhan, M., Zhang, G., Zhao, Q., Zheng, L., Zheng, X.H.,
Zhong, F.N., Zhong, W., Zhou, X., Zhu, S., Zhu, X., Smith, H.O.,
Gibbs, R.A., Myers, E.W., Rubin, G.M. and Venter, J.C.
The genome sequence of Drosophila melanogaster
Science 287 (5461), 2185-2195 (2000)
20196006
2 (bases 1 to 294218)
Adams, M.D., Celniker, S.E., Gibbs, R.A., Rubin, G.M. and Venter, C.J.
Direct Submission
Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
On Oct 9, 2000 this sequence version replaced gi:7290460.
Location/Qualifiers
1..294218
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/chromosome="X"
complement(join(3756..5473,5575..>6613))
/gene="CG6379"
/product="CT19894"
/db_xref="FLYBASE:FBgn0006379"
/db_xref="FLYBASE:FBgn0029693"
```



```

VERSION      AJ223604.1   GI:4210822
KEYWORDS
SOURCE       Pseudomonas aeruginosa.
ORGANISM     Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
              Pseudomonas.
REFERENCE    1 (bases 1 to 9581)
AUTHORS      Laraki,N., Galleni,M., Thamm,I., Riccio,M.L., Amicosante,G.,
              Frere,J.M., and Rosolini,G.M.
TITLE        Structure of In31, a blaIMP-containing Pseudomonas aeruginosa
              integron phyletically related to In5, which carries an unusual
              array of gene cassettes
JOURNAL      Antimicrob. Agents Chemother. 43 (4), 890-901 (1999)
MEDLINE      99216901
REFERENCE    2 (bases 1 to 9581)
AUTHORS      Galleni,M.G.
TITLE        Direct Submission
JOURNAL      Submitted (20-JAN-1998) Galleni M.G., Centre for Protein
              Engineering (CIP), University of Liege, B6 Sart Tilman, Liege,
              Liege B4000, Belgium
FEATURES     Location/Qualifiers
              source          1..9581
                           /organism="Pseudomonas aeruginosa"
                           /db_xref="taxon:287"
              repeat_region   117..140
                           /note="IRi"
                           /rpt_type="INVERTED"
              misc_feature    117..1467
                           /note="5' conserved element"
              gene            complement(319..1332)
                           /gene="int11"
              CDS              complement(319..1332)
                           /gene="int11"
                           /codon_start=1
                           /transl_table=11
                           /protein_id="CA11470.1"
                           /db_xref="GI:4210823"
                           /db_xref="SPTREMBL:Q9R819"
                           /translation="MKTATAPLPLRSVKVLDQLRERIRYLHYSLRTEQAYVHWVRAF
IRFHGRPATLGSSEVEAFPLSWLANERKVSVSTHQALAAALLFFYGVLCVLDLPWLQ
EIGRPPRRRLPWLTPDVRVRIIGFLEGEHRLFAQLLYGTGMRTSEGQLQKVKDLDF
DHGTTIVREGKSKDRALMPESLAPTLREOLSRARAWLKDQAGRSQVALPDALER
KYPRAGHSPWPWVFAQHTHSTDPRSGVYRRHHMDQTFQRAFKRAVEQAGITKATP
HTLRHSFATALLRSQDIDRTQDVLGHSDVSTMTIYTHVLKVGAGVRSPLDALPLPT
SER"
              mRNA            1468..2346
                           /gene="bla-imp"
              gene            1468..2346
                           /gene="bla-imp"
              CDS              1485..2225
                           /gene="bla-imp"
                           /codon_start=1
                           /transl_table=11
                           /protein_id="CA11471.1"
                           /db_xref="GI:4210824"
                           /translation="MSKLSVFIFLFCSTIATAESLPDLKIEKLDGCVVYVHTSFEEVN
GWGVPKGLVVLNAAAYLIDTPTAKDTEKLVTFVFERGYKIKGSISHPHSDSTG
GIEWLNSIPTIYASLTNELKDKGKQVATNSFGVNYVLVKNKIEFYVPGPGHTPD
NVWVLPKRKILFGGCFIKPYGLGNLDGANIEAWPKSAKLLKSKYGAALVVPVPSHSEV
GDASLLKTLLEQAVKGLNESKPKSPN"
              misc_feature    2262..2346
                           /gene="bla-imp"
                           /note="59 bp element"
              mRNA            2347..2984
                           /gene="aacA4"
              gene            2347..2984
                           /gene="aacA4"
              CDS              2371..2925
                           /gene="aacA4"
                           /codon_start=1
                           /transl_table=11
                           /protein_id="CA11472.1"
                           /db_xref="GI:4210825"

/db_xref="SPTREMBL:Q51410"
/translation="MTNSTDSVTLRLMTHEHDLAMLYEKLNRSHIVEMWGGEARPTLA
DVQEQYLPVSLAQESTPTPIAHLNGEPIGYAQSVYALCSGDGGMWEEETDPGVRGIDQL
LANASOLKGLGTKLVRALVELLFNDPEVTKIQTPDPSNLRAIRCYEKAGFERGTV
TTPDGPVAVYVQTRQAFERTSDA"
2920..2984
/gene="aacA4"
/note="59 bp element"
2985..3715
/gene="catB6"
2985..3715
/gene="catB6"
3018..3650
/gene="catB6"
/codon_start=1
/transl_table=11
/protein_id="CA11473.1"
/db_xref="GI:4210826"
/db_xref="SPTREMBL:Q9R818"
/translation="MENYFDSPEFKGKLLSEQVTNRNLIKVGYSYVSGYYHGHISFDDCA
RYLLPDRDQVDKLIIGSFCSIGSGASFIMAGNQHGRHDWVTSEPFYMOEPEAFSSST
DAFOKAGDITVGNQDVMIGSEAMIPGKIGDGAIVGSRSLVTRDVEPYTIIIGNPAKO
IKKFSDEEISLLMEMEWNWPLDKIKTAMPLLCSSDIFGLHRHWRGIAV"
3645..3715
/gene="catB6"
/note="59 bp element"
3716..4404
/gene="orf1"
3716..4404
/gene="orf1"
3727..4341
/gene="orf1"
/codon_start=1
/transl_table=11
/protein_id="CA11474.1"
/db_xref="GI:4210827"
/db_xref="SPTREMBL:Q9ZEHI"
/translation="MDVDLSIGIQTVQEALEKGLIFNEKTKTDDENKACJHILVLL
VDAQCQFDRGSWCTSVFLSHTALEEVAKEVGLYRRGCKIGKARGKDKLFNHGEKHR
MALTPVFKSRLEELALGKKECAELLKEAARHGRNRSLSYFSNENGQEVTPANVV
SQNRARELLLEAADDRLVGYTNHTGTGILEAKINEIFSVAHS"
4335..4404
/gene="orf1"
/note="59 bp element"
4405..4940
/gene="gacE2"
4405..4940
/gene="gacE2"
4507..4839
/gene="gacE2"
4507..4839
/gene="gacE2"
/codon_start=1
/transl_table=11
/protein_id="CA11475.1"
/db_xref="GI:4210828"
/db_xref="SPTREMBL:Q9ZEHO"
/translation="MKNWLFATAIIFEVIATATKLSKSGEFTRLVPSFIIVAGYAAAF
YFLSLTKSPIPGIAYAVWSGLGIVLVTIAWVLRGKLDMMGFVGVGFTIISGVAVLN
LLSKASVH"
4849..4936
/gene="gacE2"
/note="59 bp element"
4941..6961
/note="3' conserved element"
5045..5392
/gene="gacEdelta 1"
5045..5392
/gene="gacEdelta 1"
/codon_start=1
/transl_table=11
/protein_id="CA11476.1"
/db_xref="GI:4210829"

```

```
/translation="MKGWLFLVIAIIVGEVIATSKLSSEGTCKLAPSAVVIIGYIAF
YFLSLVLSKIPGVAYAVWSGLGVIIITAIWLLHQKLDANGFVGMGLIIAFLAR
SPSKSLRRPWP"
5386..6225
/gene="suli"
5386..6225
/gene="suli"
/codon_start=1
/transl_table=11
/protein_id="CAA11477.1"
/db_xref="GI:4210830"
/translation="MTVTFGLINLTEDSFFDESSRLDPAGAVTAIEMLRVGSDDVVD
GPAAGHPDAPVSPADEIRRIAPLDLSDQMHRYSIDSFOPETORYALKRGVGYLND
IGGFDPALYDPIADACDRLVVMHSAORDGIATRTCHLPEDALDEIVRFFPARVSAL
RSGVAAERLLSLDPGCGFLSPAPETSLHLV:NLQKLSALGLPLLVSVSRKSLFGAT
VGLPDLGDLPSLAAELHAIGNAGADYVRTHAPDGLRSALTSETLAKFRSDARDRL
DHA"
6353..6853
/gene="orf5"
6353..6853
/gene="orf5"
/codon_start=1
/transl_table=11
/protein_id="CAA11478.1"
/db_xref="GI:4210831"
/translation="MDSEPPNVRVACSGDIDVVRLHDAAMWSKGTPIAMDVARI
DRTFATFVLSKELVASCDSIGVCCCTLSAEDPEFPDALKGEAAYLHKLAVRTHA
GRGVSSALIERACRHAARTGGCAKRLDCHPNLRLGLYERLGTHTVDFNPGWDPTEIAE
RLELEI"
6962..7320
/misc_feature
/note="359 bp identical to In5"
complement(7311..7736)
/gene="tniB delta2"
complement(7311..7736)
/gene="tniB delta2"
/codon_start=1
/transl_table=11
/protein_id="CAA11479.1"
/db_xref="GI:4210832"
/db_xref="SPTREMBL:Q9R817"
/translation="MDEYPIIDLSHLLPAAQGLARLPADERIQRRLRDRWIGYPRAVE
ALNRLAALYAMPNQRKMPNLLVGTNNKSMIVKFRTHPASPDAQAEHPIPLVVQ
MPSPPSVTRFVYVALLAAMGAPLRPRPLPENEQLAAV"
complement(7739..9418)
/gene="tniA"
complement(7739..9418)
/gene="tniA"
/codon_start=1
/transl_table=11
/protein_id="CAA11480.1"
Query Match 45.4%; Score 23.6; DB 3; Length 9581;
Best Local Similarity 76.3%; Pred. No. 1.9e+02;
Matches 29; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
QY 15 gcggaccggctaaagccggcccttaaccacgctta 52
|||||||
Db 4370 GCTGGACGGCTAAAGCGCCCTTAACCAACGTTA 4407
RESULT 6
AC025118
LOCUS AC025118 68596 bp DNA HTG 13-JUL-2000
DEFINITION Homo sapiens chromosome 8 clone RP11-485E23 map 8, LOW-PASS
SEQUENCE SAMPLING.
ACCESSION AC025118
VERSION AC025118.1 GI:7158929
KEYWORDS HTG; HTGS_PHASE0.
SOURCE human
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS

1 (bases 1 to 68596)
Birren, B., Linton, L., Nusbaum, C. and Lander, E.
Homo sapiens chromosome 8, clone RP11-485E23
Unpublished
2 (bases 1 to 68596)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
Anderson, S., Baldwin, J., Barna, N., Bastien, V., Bida, F.,
Boguslavsky, L., Boukhgalter, B., Brown, A., Burkett, G.,
Campopiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,
Collamore, A., Cooke, P., DeArrellano, K., Dewar, K., Diaz, J. S.,
Dodge, S., Domino, M., Doyle, M., Ferreira, P., FitzHugh, W., Gage, D.,
Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,
Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Howland, J. C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
Klein, J., LaRoque, K., Lamazares, R., Landers, T., Lehoczy, J.,
Levine, R., Liu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N.,
McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R.,
Meldrum, J., Meneus, L., Mihova, T., Miranda, C., Mienda, V., Morrow, J.,
Murphy, T., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,
O'Neill, D., Oliver, T. M., Oliver, J., Peterson, K., Pierre, N.,
Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J.,
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J.,
Young, G., Zainoun, J., Zimmer, A. and Zody, M.
Direct Submission
Submitted (05-MAR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L7457
Center clone name: 485_E_23

* NOTE: This record contains 83 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be generic and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
* 1 715: contig of 715 bp in length
* 716 815: gap of 100 bp
* 816 1542: contig of 727 bp in length
* 1543 1642: gap of 100 bp
* 1643 2377: contig of 735 bp in length
* 2378 2477: gap of 100 bp
* 2478 3205: contig of 728 bp in length
* 3206 3305: gap of 100 bp
* 3306 4041: contig of 736 bp in length
* 4042 4141: gap of 100 bp
* 4142 4856: contig of 715 bp in length
* 4857 4956: gap of 100 bp
* 4957 5679: contig of 723 bp in length
* 5680 5779: gap of 100 bp
* 5780 6470: contig of 691 bp in length
* 6471 6570: gap of 100 bp
* 6571 7299: contig of 729 bp in length
* 7300 7399: gap of 100 bp
* 7400 8106: contig of 707 bp in length
* 8107 8206: gap of 100 bp
* 8207 8938: contig of 732 bp in length

```
* 8939 9038: gap of 100 bp
* 9039 9782: contig of 744 bp in length
* 9783 9882: gap of 100 bp
* 9883 10608: contig of 726 bp in length
* 10609 10708: gap of 100 bp
* 10709 11442: contig of 734 bp in length
* 11443 11542: gap of 100 bp
* 11543 12259: contig of 717 bp in length
* 12260 12359: gap of 100 bp
* 12360 13080: contig of 721 bp in length
* 13081 13180: gap of 100 bp
* 13181 13904: contig of 724 bp in length
* 13905 14004: gap of 100 bp
* 14005 14734: contig of 730 bp in length
* 14735 14834: gap of 100 bp
* 14835 15563: contig of 729 bp in length
* 15564 15663: gap of 100 bp
* 15664 16375: contig of 712 bp in length
* 16376 16475: gap of 100 bp
* 16476 17198: contig of 724 bp in length
* 17200 17299: gap of 100 bp
* 17300 18031: contig of 732 bp in length
* 18032 18131: gap of 100 bp
* 18132 18874: contig of 743 bp in length
* 18875 18974: gap of 100 bp
* 18975 19693: contig of 725 bp in length
* 19700 19799: gap of 100 bp
* 19800 20524: contig of 725 bp in length
* 20525 20624: gap of 100 bp
* 20625 21333: contig of 709 bp in length
* 21334 21433: gap of 100 bp
* 21434 22172: contig of 739 bp in length
* 22173 22272: gap of 100 bp
* 22273 23005: contig of 733 bp in length
* 23006 23105: gap of 100 bp
* 23106 23848: contig of 743 bp in length
* 23849 23948: gap of 100 bp
* 23949 24676: contig of 728 bp in length
* 24677 24776: gap of 100 bp
* 24777 25494: contig of 718 bp in length
* 25495 25594: gap of 100 bp
* 25595 26332: contig of 738 bp in length
* 26333 26432: gap of 100 bp
* 26433 27173: contig of 741 bp in length
* 27174 27273: gap of 100 bp
* 27274 28036: contig of 763 bp in length
* 28037 28136: gap of 100 bp
* 28137 28855: contig of 719 bp in length
* 28856 28955: gap of 100 bp
* 28956 29688: contig of 733 bp in length
* 29689 29788: gap of 100 bp
* 29789 30535: contig of 747 bp in length
* 30536 30635: gap of 100 bp
* 30636 31368: contig of 733 bp in length
* 31369 31468: gap of 100 bp
* 31469 32201: contig of 733 bp in length
* 32202 32301: gap of 100 bp
* 32302 33012: contig of 711 bp in length
* 33013 33112: gap of 100 bp
* 33113 33853: contig of 741 bp in length
* 33854 33953: gap of 100 bp
* 33954 34674: contig of 721 bp in length
* 34675 34774: gap of 100 bp
* 34775 35496: contig of 722 bp in length
* 35497 35596: gap of 100 bp
* 35597 36334: contig of 738 bp in length
* 36335 36434: gap of 100 bp
* 36435 37172: contig of 738 bp in length
* 37173 37272: gap of 100 bp
* 37273 37998: contig of 726 bp in length
* 37999 38098: gap of 100 bp
* 38099 38820: contig of 722 bp in length
* 38821 38920: gap of 100 bp
```

```
* 38921 39621: contig of 701 bp in length
* 39622 39721: gap of 100 bp
* 39722 40427: contig of 706 bp in length
* 40428 40527: gap of 100 bp
* 40528 41260: contig of 733 bp in length
* 41261 41360: gap of 100 bp
* 41361 42100: contig of 740 bp in length
* 42101 42200: gap of 100 bp
* 42201 42932: contig of 732 bp in length
* 42933 43032: gap of 100 bp
* 43033 43747: contig of 715 bp in length
* 43748 43847: gap of 100 bp
* 43848 44577: contig of 730 bp in length
* 44578 44677: gap of 100 bp
* 44678 45392: contig of 715 bp in length
* 45393 45492: gap of 100 bp
* 45493 46209: contig of 717 bp in length
* 46210 46309: gap of 100 bp
* 46310 47028: contig of 719 bp in length
* 47029 47128: gap of 100 bp
* 47129 47852: contig of 724 bp in length
* 47853 47952: gap of 100 bp
* 47953 48674: contig of 722 bp in length
* 48675 48774: gap of 100 bp
* 48775 49504: contig of 730 bp in length
* 49505 49604: gap of 100 bp
* 49605 50333: contig of 729 bp in length
* 50334 50433: gap of 100 bp
* 50434 51170: contig of 737 bp in length
* 51171 51270: gap of 100 bp
* 51271 52001: contig of 731 bp in length
* 52002 52101: gap of 100 bp
* 52102 52838: contig of 737 bp in length
* 52839 52938: gap of 100 bp
* 52939 53671: contig of 733 bp in length
* 53672 53771: gap of 100 bp
* 53772 54533: contig of 762 bp in length
* 54534 54633: gap of 100 bp
* 54634 55354: contig of 721 bp in length
* 55355 55454: gap of 100 bp
* 55455 56183: contig of 729 bp in length
* 56184 56283: gap of 100 bp
* 56284 57008: contig of 725 bp in length
* 57009 57108: gap of 100 bp
```

Query Match 44.6%; Score 23.2; DB 69; Length 68596;
Best Local Similarity 70.5%; Pred. No. 1.7e+02;
Matches 31; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

```
Qy 9 ctcaactcgaggacgcgctaaagccgcgccttaaccacgta 52
||||| ||| ||||| || ||||| || |||||
Db 54835 CTCACGTGTTCCGGCTACAGTTGCCCTTGATGAGTGTTA 54878
```

RESULT 7

CNS01DWP

LOCUS

DEFINITION

Human chromosome 14 DNA sequence *** IN PROGRESS *** BAC R-102C24

complete sequence.

AL137818

VERSION

HTG; HTGS_DRAFT.

SOURCE

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 181052)

Genoscope.

Direct Submission

Submitted (23-MAY-2000) to the EMBL/GenBank/DBJ databases

On Feb 20, 2000 this sequence version replaced gi:6982193.

IMP: TANT: This sequence is unfinished and does not necessarily

PRI 25-MAY-2000

represent the correct sequence. Work on the sequence is in progress and the release of this data is based on the understanding that the sequence may change as work continue. The sequence may be contaminated with foreign sequence from E.coli, yeast, vector, phage, etc. . . even if efforts are made to eliminate these contaminating sequences. The following BAC sequence is oriented from the T7 to the SP6 end.

Upstream BAC (overlapping the T7 end) : R-317N8
Downstream BAC (overlapping the SP6 end) : R-355C3

Overall quality chart :

Range : bases

0 : bases

1 - 9 : 93

10 - 19 : 320

20 - 29 : 1291

30 - 39 : 3128

40 - 49 : 5327

50 - 59 : 12776

60 - 69 : 30990

70 - 79 : 58042

80 - 89 : 42197

90 - 99 : 26888

Percentage of bases with a quality value >= 40 : 97 %.

Location/Qualifiers

1..181052

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="14"

/clone_lib="RPC1-11"

/clone="R-102C24"

BASE COUNT 52464 a 32038 c 33821 g 62729 t

ORIGIN

Query Match 44.6%; Score 23.2; DB 91; Length 181052;
Best Local Similarity 70.5%; Pred. No. 1.4e+02;
Matches 31; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 9 ctaactgcggagccgctaaagccggcccttaaccaaagctta 52

Db 116656 CTCACGTGTGTCGGCTACAGTTTGCCTTGATGAGTGTTA 116699

RESULT 8

BSU11039

LOCUS

DEFINITION

BSU11039 33615 bp DNA BCT 06-JAN-1996

Bacillus subtilis w168 polyketide synthase (pkx and pksofx6)

genes, complete cds.

ACCESSION

U11039 M97902

VERSION

U11039.1 GI:602656

KEYWORDS

SOURCE

ORGANISM

Bacillus subtilis.

Bacillus subtilis

Bacteria; Firmicutes; Bacillus/Clostridium group;

Bacillus/Staphylococcus group; Bacillus.

1 (bases 2000 to 33615)

Scotti,C., Platti,M., Cuzzoni,A., Perani,P., Tognoni,A., Grandi,G.,

Galizzi,A.M. and Albertini,A.M.

A Bacillus subtilis large ORF coding for a polypeptide highly

similar to polyketide synthases

Gene 130 (1), 65-71 (1993)

93345824

2 (bases 1 to 20000)

Albertini,A.M., Caramori,T., Scoffone,F., Scotti,C. and Galizzi,A.

Sequence around the 159 degree region of the Bacillus subtilis

genome: the pkx locus spans 33.6 kb

Microbiology 141 (Pt 2), 299-309 (1995)

95219083

JOURNAL

MEDLINE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

Direct Submission

JOURNAL

COMMENT

FEATURES

source

1..33615

/organism="Bacillus subtilis"

/strain="168"

/db_xref="taxon:1423"

323..369

785..789

795..2045

/note="pksofx1: similar to Rhizobium leguminosarum

nodulation protein E, Swiss-Prot Accession Number P04684,

and to E. coli 3-oxoacyl-(acyl-carrier protein) synthase

I, Swiss-Prot Accession Number P14926"

/codon_start=1

/transl_table=11

/product="unknown"

/protein_id="AA085139.1"

/db_xref="GI:602657"

/translation="MTKCNLPVVVTVGTASIGOGKEDFASLLSGRHAFDVMKRS

GRKDSRFIGAELASLSYDRLSKMLKASPSRRAALVLTAEWEAEADDDASRI

GLVGSNPFQQRNFYQDRSGFISPAYGLSFMDSDLGICTDQFGITGLATV

GGASAGQQLAVIHAIQVLSGEVDTICIALGMLDSLMECEALSGAMGTDKYADE

PENACRPFDQNRDGFYIGSCGALVIERKETAIRGLKPYAALSGWSIKLDGNNRDP

SLEGEITHVQKALERARLLPEDIDYINPHGTGTFIGDETELKALRACRLSHAYINATK

SITGHLSAAGIVEIISVLLQMKKSAHPSRNLDHPIDDDSFHWVNEKSIYRIKNALS

LSMGFGMNTAVCQNIKKGES"

1961..1989

-35_signal

-10_signal

2059..2063

/note="sigma H consensus"

2253..2258

2270..3307

/note="pksofx2: similar to hamster

hydroxymethylglutaryl-CoA synthase, Swiss-Prot Accession

Number P13704"

/codon_start=1

/transl_table=11

/product="unknown"

/protein_id="AA085140.1"

/db_xref="GI:528991"

/translation="NLITCSSESGIDFGKSLSTYIHEYLGILNRNCRLEFVKQACYSCTA

GFQMAVNFILSQTSPGAKALVIASDISRFLIAEGGDSLEDWSYAEPSAGAVAVLV

GENPEVFOIDPGANGYGVGVNDTCRPIPDSEAGSDLSMSYLDCCETFEYOKRV

PGANYQDTFOYLAYHTPEGGVKGARHTMRKVAVKTSIETDFTLVKPGNLTCOR

VGNMGALFLALASTIQDREDTPRIGCFSTYSGCCSEFYSGITTPQGGQRQRTFG

IEKHLDRIYQLSMEEYELLFKSGMVRFGRNVKLDFFEMIPGIMOSTQEKPLFLEEI

SEFHRKYRWS"

3285..3290

3295..4074

/note="pksofx3: similar to Rhodobacter capsulatus

enoyl-CoA hydratase Swiss-Prot Accession Number P24162"

/codon_start=1

/transl_table=11

/product="unknown"

/protein_id="AA085141.1"

/db_xref="GI:528992"

/translation="MDLVYQTIKVRFOASVCYITFHRPEANNTINDTJIECLQVLN

QCESTVTVVVLEGLPEVFCFGADFOEYQEMKGRKQASOEPLYDLMMKLQTPYV

TISHVRGNAGGLGFVSATIDATADQATASFSLSELLFGLYPACVLPFLIRIRGQKA

HYMTLMTKPIISVQEAASEMGLIDAFDAESDVLRLKLLRLRLNKKGIAHAKQFMSSLD

HOVSRKATALTANQDMFSDPQNQNGIIRYVETGQFPWEDQ"

4103..4108

4114..4863

/note="pksofx4: similar to Rhodobacter capsulatus

enoyl-CoA hydratase Swiss-Prot Accession Number P24162"

/codon_start=1

/transl_table=11

/product="unknown"

/protein_id="AA085142.1"

/db_xref="GI:528993"

Baker, J., Baldwin, J., Barna, N., Beckerly, R., Benn, J., Brown, A., Castle, A., Cerny, J., Collangelo, M., Collins, S., Collymore, A., Cooke, P., DeArelano, K., Depayre, E., Devon, K., Dewar, K., Donelan, L., Doyle, M., Ferrelira, P., FitzHugh, W., Forrest, C., Funke, B., Gage, D., Galadan, J., Gardyna, S., Gilbert, D., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J. C., Jones, C., Kann, L., Karatas, A., Lehoczy, J., Lieu, C., Locke, K., Macdonald, P., Marquis, N., McEwan, P., McGurk, A., McKernan, K., McDonald, P., Meldrim, J., Molla, M., Morris, W., Morrow, J., Mychaleckyj, J., Naylor, J., Niloff, M., O'Connor, T., O'Donnell, P., Pavlin, B., Peterson, K., Pollara, V., Riley, R., Roberts, D., Roy, A., Severy, P., Stange-Thomann, N., Stojanovic, N., Stone, C., Subramanian, A., Testaye, S., Torruella-Miller, I., Vassiliev, H., Vo, A., Wagner, A., Wheeler, J., Wu, X., Wyman, D., Ye, W. J. and Zody, M.

Direct Submission

Submitted (17-APR-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

On Apr 16, 1999 this sequence version replaced gi:4580413.

All repeats were identified using RepeatMasker: Smit, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>.

FEATURES

source

1. .120766

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="hRPC 971_F_3"

/clone_lib="Peter de Jong/ human PAC library"

/map="17"

/chromosome="17"

1285..1324

/rpt_family="CGTG)n"

complement(1764..1935)

/rpt_family="MER91A"

8156..8350

/rpt_family="MIR"

11510..11540

/rpt_family="(T)n"

11876..11927

/rpt_family="L2"

complement(12864..12966)

/rpt_family="AluSg"

complement(13019..13043)

/rpt_family="MIR"

13044..13334

/rpt_family="AluJo"

complement(13335..13393)

/rpt_family="MIR"

13430..13699

/rpt_family="AluSg"

complement(13704..13872)

/rpt_family="MER91A"

13954..13992

/rpt_family="MER65B"

complement(14240..14426)

/rpt_family="MERSA"

15195..15325

/rpt_family="MIR"

15946..15960

/rpt_family="Charlie4"

complement(15961..16254)

/rpt_family="AluSg"

16255..16285

/rpt_family="Charlie4"

16286..16588

/rpt_family="AluSg"

16589..16661

/rpt_family="Charlie4"

complement(16698..16900)

/rpt_family="MIR"

17327..17631

/rpt_family="AluSg"

17635..17928

/rpt_family="AluSc"

repeat_region

complement(17932..17987)

/rpt_family="MIR"

repeat_region

complement(18019..18188)

/rpt_family="MER91A"

18539..18655

/rpt_family="MIR"

repeat_region

complement(18859..19170)

/rpt_family="AluY"

19182..19205

/rpt_family="AT_rich"

19206..19505

/rpt_family="AluSc"

complement(19218..19229)

/note="Single-stranded terminator coverage."

19506..19524

/rpt_family="AT_rich"

complement(19525..19837)

/rpt_family="AluSg"

19838..19858

/rpt_family="AT_rich"

20517..20822

/rpt_family="AluY"

complement(21146..21433)

/rpt_family="AluSp"

complement(21532..21835)

/rpt_family="AluSx"

23272..23500

/rpt_family="AluSg"

23518..23827

/rpt_family="AluSg"

24479..24589

/rpt_family="C-rich"

24984..25165

/rpt_family="MIR"

25207..25471

/rpt_family="AluJo"

25474..25560

/rpt_family="(TA)n"

complement(25637..25933)

/rpt_family="AluY"

complement(25940..26250)

/rpt_family="AluSx"

26573..26637

/rpt_family="(CAG)n"

26800..26939

/rpt_family="(CAG)n"

complement(29224..29299)

/rpt_family="MER33"

29300..29606

/rpt_family="AluY"

complement(29607..29624)

/rpt_family="MER33"

29625..29908

/rpt_family="AluSg"

complement(29909..29969)

/rpt_family="MER33"

complement(29997..30045)

/rpt_family="AluJo"

complement(30046..30333)

/rpt_family="AluSx"

30222..30254

/note="Single-stranded coverage."

complement(30338..30515)

/rpt_family="AluY"

30338..30358

/note="Possibly 21 T's, possibly 20 T's."

30366

/note="Probably A, possibly G."

complement(30516..30828)

/rpt_family="AluSg"

complement(30829..30964)

/rpt_family="AluY"

complement(30965..31058)

repeat_region

repeat_region	/rpt_family="AluJo"									
repeat_region	complement(31059..31214)									
repeat_region	/rpt_family="MER33"									
repeat_region	31829..31896									
repeat_region	/rpt_family="(CAG)n"									
repeat_region	31991..32280									
repeat_region	/rpt_family="AluSg"									
repeat_region	32281..32442									
repeat_region	/rpt_family="AluSg/x"									
repeat_region	32753..32794									
repeat_region	/rpt_family="(GGA)n"									
repeat_region	33138..33190									
repeat_region	/rpt_family="(TG)n"									
repeat_region	33285..33315									
repeat_region	/rpt_family="(CCCCCG)n"									
repeat_region	34350..34489									
repeat_region	/rpt_family="AluJo"									
repeat_region	34490..34784									
repeat_region	/rpt_family="AluX"									
repeat_region	34785..34963									
repeat_region	/rpt_family="AluJo"									
Query Match	44.2%: Score 23; DB 85; Length 120766;									
Best Local Similarity	74.4%: Pred. No. 1.Be+02;									
Matches	29: Conservative 0; Mismatches 10; Indels 0; Gaps 0;									
QY	1 tcgcttcgctcactcgggacgcgctaagcgcgcctc 39									
Db	53624 TCGCGCGCTCACTGGCGCGCGGCTTCGCCCGCTGCT 53662									
RESULT 10										
BSUB0009										
LOCUS										
DEFINITION	BSUB0009 208780 bp DNA BCT 26-NOV-1997									
ACCESSION	299112 AL009126									
VERSION	299112.1 GI:2633902									
KEYWORDS										
SOURCE	Bacillus subtilis.									
ORGANISM	Bacillus subtilis									
REFERENCE	Bacteria: Firmicutes: Bacillus/Clostridium group; Bacillus.									
AUTHORS	1 (bases 1 to 208780)									
	Kunst, F., Ogasawara, N., Moszer, I., Albertini, A.M., Alloni, G., Azevedo, V., Bertero, M.G., Bessieres, P., Bolotin, A., Borchert, S., Borriss, R., Boursier, L., Brans, A., Braun, M., Brignell, S.C., Bron, S., Brouillet, S., Bruschi, C.V., Caldwell, B., Capuano, V., Carter, N.M., Choi, S.K., Codani, J.J., Connerton, I.F., Cummings, N.J., Daniel, R.A., Denizot, F., Devine, K.M., Dusterhoft, A., Ehrlich, S.D., Emmerston, P.J., Entian, K.D., Errington, J., Fabret, C., Ferrari, E., Foulger, D., Fritz, C., Fujita, M., Fujita, Y., Fuma, S., Gallizzi, A., Galleron, N., Ghim, S.Y., Glaser, P., Goffeau, A., Gollightly, E.J., Grandi, G., Guiseppe, G., Guy, B.J., Haga, K., Haiech, J., Harwood, C.R., Henaut, A., Hilbert, H., Holsappel, S., Hosono, S., Hulio, M.F., Itaya, M., Jones, L., Joris, B., Karamata, D., Kasahara, Y., Klabir-Blanchard, G., Krogh, S., Kumano, M., Kurita, K., Lapidus, A., Koningsstein, G., Lauber, J., Lazarevic, V., Lee, S.M., Levine, A., Liu, H., Masuda, S., Maue, C., Medigue, C., Medina, N., Mellado, R.P., Mizuno, K., Moestl, D., Nakai, S., Noback, M., Noone, D., O'Reilly, M., Portetelie, D., Porwollik, S., Prescott, A.M., Parro, V., Pohl, T.M., Purnelle, B., Rapoport, G., Rey, M., Reynolds, S., Rieger, M., Scanlan, E., Schleich, S., Schroeter, R., Scoffone, F., Sekiguchi, J., Sekowska, A., Seror, S.J., Serron, P., Shin, B.S., Soldo, B., Sorokin, A., Tacconi, E., Takagi, T., Takahashi, H., Takemaru, K., Takeuchi, M., Tanakoshi, A., Tanaka, T., Terpstra, P., Tognoni, A., Totsato, V., Uchiyama, S., Vandenbol, M., Vannier, F., Vassarotti, A., Viari, A., Wambutt, R., Wedler, E., Wedler, H., Weitzenecker, T., Winters, P., Wipat, A., Yamamoto, H., Yamane, K., Yasumoto, K., Yata, K., Yoshida, K., Yoshikawa, H.F., Zumstein, E., Yoshikawa, H. and									

```

NTDPLEAIEQLDDSFRLDKMRLIPYRGVGQONQFLLCVKPDHVTIMTKEMISADKC
5746..6465
/ene="sigE"
CDS
5746..6465
/ene="sigE"
/function="early mother cell-specific gene expression"
/codon_start=1
/transl_table=11
/product="RNA polymerase sporulation-specific sigma-29
factor (sigma-E)"
/protein_id="CAB13406.1"
/db_xref="GI:2633905"
/db_xref="SWISS-PROT:P06222"
/translation="MKLKLRLTHLWYLLMKLGLKSDVYVIGSEALPPPLSKDEE
QVLLMKLPNGDQAARILIERILRLVYIARFENTGINIEDLISIGFTGLIKAVNTF
NPEKIKIATYASRCILENIMLYRNKIRSEVSYFELNIDWDGNELLSDVLGTD
DDITTKDEANVDYDKLLKALQELNREKQIMELRFLVGBEEKTKQDVADMMGISQS
YISLKEKRIIKRLKEFNKV"
6605..7387
/ene="sigG"
CDS
6605..7387
/ene="sigG"
/function="late foresore-specific gene expression"
/ene="alternate gene name: spoIIIG"
/codon_start=1
/transl_table=11
/product="RNA polymerase sporulation-specific sigma factor
(sigma-G)"
/protein_id="CAB13407.1"
/db_xref="GI:2633906"
/db_xref="SWISS-PROT:P19940"
/translation="MSRNKVEICGVDTSKLPVLKNEMRKLFRLQLODGGDSAREKLV
NGNLRVLSVIORENRGEYVDDLPQVCGICIGMKSIDNFDLSHNKVFSTYAVPMIIGE
IRYLRDNPVRSRLDIAYKALQVRLISETSKEPTAEDTAKVLEVPREEHIVFA
LDAIDQVSLSEPIYNDGGDPIIYHQDQISDERNTDSOWIEELALKEGMRRLNDRKMI
LKRFFQKQTQMEVAEIGISQAQVSRLEKAAIKQMNKNIHQ"
7535..8496
/ene="ylmA"
CDS
7535..8496
/ene="ylmA"
/function="unknown"
/ene="similar to ABC transporter (ATP-binding protein)"
/codon_start=1
/transl_table=11
/protein_id="CAB13408.1"
/db_xref="GI:2633907"
/db_xref="SPTREMBL:O31723"
/translation="MILQLDNVSLKRNCKWILKDIHWKVEKENWVLYGLNAGKTAL
LNMCSYFPTSGEMRVLGHFGKTELGELKRRKIGLVSAALQKLYPADSAFETALS
GAYASIGLYETPSKTEKAKGLLEDLGAIEYADRYETLSQGEKORALIALRAMDAP
ELLIDPEVTGLDFTIAREKLLDTITYIANKENAPSILYVTHAEIILPVFDKALLKQ
GEVFGSGIKEMLTQILSAFDTPIHVLWNQDRPFLTRAEPITNA"
8472..8496
/ene="ylmB"
terminator
8531..9811
/ene="ylmB"
CDS
8531..9811
/ene="ylmB"
/function="unknown"
/ene="similar to acetylornithine deacetylase"
/codon_start=1
/transl_table=11
/protein_id="CAB13409.1"
/db_xref="GI:2633908"
/db_xref="SPTREMBL:O31724"
/translation="MDQIYSLQKKVEHKEELIQAKTLISVQTAPPARNTEGIGS
WIAGYLNELGFSIDKWDYVPGDENVGRLKGTQDSADYISLIINGVDVDAEVEDEENK
HDPHPTEKNGLLIGRASDMKGMACVLFVAKVLIAREASIELPGDLTLQSVIGEEVGE
AGTCCCKRGTHADFAIVADTSDMHIOGGVITGWIEIKSSQTFDHDTRRNMIHAGG
GTGFAECERMAKTIAGLGERLHWSIMKSPYGPKPGTINPAPIEGGRHAAFIADG
CRLWTVHFPNETHDQVAAEIEDYVNRLLSDSDIWLRENRPVFKWGGSSMIEDRGEIF

```

```

PALEVDPPGHPGVLTALTASHQKVKRECPIDVDSQSVTDGGWLYDAGIPCVIYGPCLDHN
AHSVNEKVSIEQLVEYTKIILDPIISWCSRKKEQ"
9894..10139
/ene="ylmC"
CDS
9894..10139
/ene="ylmC"
/function="unknown"
/ene="similar to hypothetical proteins"
/codon_start=1
/transl_table=11
/protein_id="CAB13410.1"
/db_xref="GI:2633909"
/db_xref="SPTREMBL:O31725"
/translation="MISISEFQVKDVVNVSNKKLGSIGDIDINVTGKIQAIIILGGN
Query Match 44.2%; Score 23; DB 2; Length 208780;
Best Local Similarity 74.4%; Pred. No. 1.6e+02;
Matches 29; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
QY 13 ctgcgggacgcgctaaagccgcgccttaacacaaacgtt 51
|||||
Db 199854 CTGGGGGACCCGATCGACTGCGCGCTTTAACAAACGTT 199892

RESULT 11
AC023855 218864 bp DNA HTG 10-JAN-2001
LOCUS Homo sapiens chromosome 17 clone RP11-546M21 map 17, WORKING DRAFT
DEFINITION AC023855
AC023855
HTG: HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 218864)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
JOURNAL Homo sapiens chromosome 17, clone RP11-546M21
TITLE Unpublished
JOURNAL
REFERENCE 2 (bases 1 to 218864)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beda,F., Boguslavsky,L.,
Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
Choepe,Y., Colangelo,M., Collins,S., Collamore,A., Cooke,P.,
Dearellano,K., Dewar,K., Dodge,S., Domino,M., Doyle,M.,
Fenster,J., Ferreira,P., FitzHugh,W., Forrest,C., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., Landers,T., Largocque,K., Lehoczy,J., Levine,R.,
Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M.,
McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrum,J.,
Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J., Nayler,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Olivier,T.M.,
Peterson,K., Pierre,N., Pisan,C., Pollara,V., Raymond,C.,
Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,
Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A.,
Travers,M., Triglio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,E.,
Wu,X., Wyman,D., Ye,W.J., Young,G., Zalnoun,J., Zimmer,A. and
Zody,M.
Direct Submission
TITLE Submitted (18-FEB-2000) Whitehead Institute/MIT Center for Genome
JOURNAL Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On Jan 10, 2001 this sequence version replaced gi:8954143.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Web code: WBIR
Web site: http://www-seq.wi.mit.edu

```



```

VERSION      AF033487.1  GI:2668727
KEYWORDS
SOURCE       Eupenicillium tularensis.
ORGANISM     Eupenicillium tularensis.
REFERENCE    Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
AUTHORS      Eutiales; Trichosporaceae; Eupenicillium.
TITLE        1 (bases 1 to 1142)
JOURNAL      Peterson,S.W.
              Phylogenetic analysis of Penicillium species based on ITS and
              LSU-rDNA nucleotide sequences
              (in) Samson,R.A. and Pitt,J.I. (Eds.);
              INTEGRATION OF MODERN TAXONOMIC METHODS FOR PENICILLIUM AND
              ASPERGILLUS CLASSIFICATION: 163-178;
              HARWOOD Academic Publishers, The Netherlands (2000)
REFERENCE    2 (bases 1 to 1142)
AUTHORS      Peterson,S.W.
TITLE        Phylogenetic analysis of Penicillium species based on ITS and
              LSU-rDNA nucleotide sequences
JOURNAL      Unpublished
REFERENCE    3 (bases 1 to 1142)
AUTHORS      Peterson,S.W.
TITLE        Direct Submission
JOURNAL      Submitted (07-NOV-1997) Microbial Properties Research, National
              Center for Agricultural Utilization Research, 1815 N. University
              St., Peoria, IL 61604, USA
FEATURES     source
              Location/Qualifiers
              1..1142
                /organism="Eupenicillium tularensis"
                /strain="NRRL 5273"
                /db_xref="taxon:36626"
              misc_RNA
                1..184
                  /product="internal transcribed spacer 1"
                  185..341
                    /product="5.8S"
              misc_RNA
                342..506
                  /product="internal transcribed spacer 2"
                  507..>1142
                    /product="28S"
BASE COUNT   255 a 306 c 347 g 234 t
ORIGIN
Query Match      43.5%; Score 22.6; DB 12; Length 1142;
Best Local Similarity 68.9%; Pred. No. 6.9e+02;
Matches 31; Conservative 0; Mismatches 14; Indels 0; Gaps 0;
QY 7  cgcctactgcggacgcggttaagcggcccttaaccacagctt 51
    | | | | | | | | | | | | | | | | | | | | | | |
Db 454  CCCGCTCGAGCGCGCGCGCGCGCGCGCGCGGATCAACCAACTTT 498

RESULT 15
AC084177      AC084177  45257 bp  DNA      HTG      14-OCT-2000
LOCUS          Homo sapiens chromosome 16 clone RP11-616M22 map 16, LOW-PASS
DEFINITION     SEQUENCE SAMPLING.
ACCESSION      AC084177
VERSION        AC084177.1  GI:10801405
KEYWORDS       HTG; HTGS_PHASE0.
SOURCE         human.
ORGANISM       Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
              Anderson,S., Barna,N., Bastien,V., Beda,F., Boguslavskiy,L.,
              Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
              Choepel,Y., Colangelo,M., Collins,S., Collumore,A., Cooke,P.,
              DeArelano,K., Dewar,K., Diaz,J.S., Dodge,S., Ferreira,P.,
              2 (bases 1 to 45257)
              Homo sapiens chromosome 16, clone RP11-616M22
              Unpublished
              Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
              Anderson,S., Barna,N., Bastien,V., Beda,F., Boguslavskiy,L.,
              Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
              Choepel,Y., Colangelo,M., Collins,S., Collumore,A., Cooke,P.,
              DeArelano,K., Dewar,K., Diaz,J.S., Dodge,S., Ferreira,P.,

```

TITLE JOURNAL COMMENT

```

Fitzhugh,W., Cage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M.,
Graham,L., Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L.,
Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., LaRocque,K.,
Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Lieu,C., Liu,G.,
Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McKernan,K.,
McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T.,
O'Donnell,P., O'Neill,D., Ollivar,T.M., Oliver,J., Peterson,K.,
Pierre,N., Pisani,C., Pollara,V., Raymond,C., Rieback,M., Riley,R.,
Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P.,
Sougez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,
Tirrell,A., Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A.,
Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
Zimmer,A. and Zody,M.
Direct Submission
Submitted (14-OCT-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit,A.F.A. & Green,P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L11412
Center clone name: 616_M_22
-----
* NOTE: This record contains 55 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
* 1 756: contig of 756 bp in length
* 757 856: gap of 100 bp
* 857 1616: contig of 760 bp in length
* 1617 1716: gap of 100 bp
* 1717 2434: contig of 718 bp in length
* 2435 2534: gap of 100 bp
* 2535 3271: contig of 737 bp in length
* 3272 3371: gap of 100 bp
* 3372 4120: contig of 749 bp in length
* 4121 4220: gap of 100 bp
* 4221 4948: contig of 728 bp in length
* 4949 5048: gap of 100 bp
* 5049 5765: contig of 717 bp in length
* 5766 5865: gap of 100 bp
* 5866 6564: contig of 699 bp in length
* 6565 6664: gap of 100 bp
* 6665 7376: contig of 712 bp in length
* 7377 7476: gap of 100 bp
* 7477 8197: contig of 721 bp in length
* 8198 8297: gap of 100 bp
* 8298 9048: contig of 751 bp in length
* 9049 9148: gap of 100 bp
* 9149 9872: contig of 724 bp in length
* 9873 9972: gap of 100 bp
* 9973 10697: contig of 725 bp in length
* 10698 10797: gap of 100 bp
* 10798 11525: contig of 728 bp in length
* 11526 11625: gap of 100 bp
* 11626 12347: contig of 722 bp in length
* 12348 12447: gap of 100 bp
* 12448 13188: contig of 741 bp in length
* 13189 13288: gap of 100 bp

```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw 4.4.0.1

Run on: August 2, 2001, 16:55:10 ; Search time 1874.14 Seconds
(without alignments)
388.375 Million cell updates/sec

Title: US-09-701-626A-78
Perfect score: 77
Sequence: 1 ccttaacaaagtgtcaagc.....cccttaaccaaacgtagag 77

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues 20456230
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*

1:	gb_est1:*
2:	gb_est2:*
3:	gb_est3:*
4:	gb_est4:*
5:	gb_est5:*
6:	gb_est6:*
7:	gb_est7:*
8:	gb_est8:*
9:	gb_est9:*
10:	gb_est10:*
11:	gb_est11:*
12:	gb_est12:*
13:	gb_est13:*
14:	gb_est14:*
15:	gb_est15:*
16:	gb_est16:*
17:	gb_est17:*
18:	gb_est18:*
19:	gb_est19:*
20:	gb_est20:*
21:	gb_est21:*
22:	gb_est22:*
23:	gb_est23:*
24:	gb_est24:*
25:	gb_est33:*
26:	gb_est34:*
27:	gb_est35:*
28:	gb_est36:*
29:	gb_est37:*
30:	gb_est38:*
31:	gb_est39:*
32:	gb_est40:*
33:	em_estba:*
34:	em_estfun:*
35:	em_esthum1:*
36:	em_esthum2:*
37:	em_esthum3:*
38:	em_esthum4:*
39:	em_esthum5:*
40:	em_esthum6:*
41:	em_esthum7:*
42:	em_esthum8:*
43:	em_esthum9:*

44:	em_esthum10:*
45:	em_esthum11:*
46:	em_esthum12:*
47:	em_esthum13:*
48:	em_esthum14:*
49:	em_esthum15:*
50:	em_esthum16:*
51:	em_esthum17:*
52:	em_esthum18:*
53:	em_esthum19:*
54:	em_esthum20:*
55:	em_esthum21:*
56:	em_esthum22:*
57:	em_esthum23:*
58:	em_esthum24:*
59:	em_esthum25:*
60:	em_esthum26:*
61:	em_esthum27:*
62:	em_esthum28:*
63:	em_estin1:*
64:	em_estin2:*
65:	em_estin3:*
66:	em_estin4:*
67:	em_estin5:*
68:	em_estom1:*
69:	em_estom2:*
70:	em_estov1:*
71:	em_estov2:*
72:	em_estpl1:*
73:	em_estpl2:*
74:	em_estpl3:*
75:	em_estpl4:*
76:	em_estpl5:*
77:	em_estpl6:*
78:	em_estpl7:*
79:	em_estpl8:*
80:	em_estpl9:*
81:	em_estpl10:*
82:	em_estro1:*
83:	em_estro2:*
84:	em_estro3:*
85:	em_estro4:*
86:	em_estro5:*
87:	em_estro6:*
88:	em_estro7:*
89:	em_estro8:*
90:	em_estro9:*
91:	em_estro10:*
92:	em_estro11:*
93:	em_estro12:*
94:	em_estro13:*
95:	em_estro14:*
96:	em_estro15:*
97:	em_estro16:*
98:	em_estro17:*
99:	em_estro18:*
100:	em_estro19:*
101:	em_estro20:*
102:	gb_est25:*
103:	gb_est26:*
104:	gb_est27:*
105:	gb_est28:*
106:	gb_est29:*
107:	gb_est30:*
108:	gb_est31:*
109:	gb_est32:*
110:	gb_est41:*
111:	gb_est42:*
112:	gb_est43:*
113:	gb_est44:*
114:	gb_est45:*
115:	gb_est46:*
116:	gb_est47:*

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*

190: gb_est121:*
191: gb_est122:*
192: gb_est123:*
193: gb_est124:*
194: gb_est125:*
195: gb_est126:*
196: gb_est127:*
197: gb_est128:*
198: gb_est129:*
199: gb_est130:*
200: gb_est131:*
201: gb_est132:*
202: gb_est133:*
203: gb_est134:*
204: gb_est135:*
205: gb_est136:*
206: gb_est137:*
207: gb_est138:*
208: gb_est139:*
209: gb_est140:*
210: gb_est141:*
211: gb_est142:*
212: gb_est143:*
213: gb_est144:*
214: gb_est145:*
215: gb_est146:*
216: gb_est147:*
217: gb_est148:*
218: gb_est149:*
219: gb_est150:*
220: gb_est151:*
221: gb_est152:*
222: gb_est153:*
223: gb_est154:*
224: gb_est155:*
225: gb_est156:*
226: gb_est157:*
227: gb_est158:*
228: gb_est159:*
229: gb_est160:*
230: gb_est161:*
231: gb_est162:*
232: gb_est163:*
233: gb_est164:*
234: gb_est165:*
235: gb_est166:*
236: gb_est167:*
237: gb_est168:*
238: gb_est169:*
239: gb_est170:*
240: gb_est171:*
241: gb_est172:*
242: gb_est173:*
243: gb_est174:*
244: gb_est175:*
245: gb_est176:*
246: gb_est177:*
247: gb_est178:*
248: gb_est179:*
249: gb_est180:*
250: gb_est181:*
251: gb_est182:*
252: gb_est183:*
253: gb_est184:*
254: gb_est185:*
255: gb_est186:*
256: gb_est187:*
257: gb_est188:*
258: gb_est189:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES												
Result No.	Score	Query Match	Length	DB	ID	Description						
C 1	28.6	37.1	510	258	TA382G08P	AL498114 T. brucei						
C 2	27.6	35.8	610	138	BE620662	BE620662 601483644						
C 3	27	35.1	536	105	AL524576	AL524576 AL524576						
C 4	26	33.8	675	3	AA203709	AA203709 zx52d10.r						
C 5	26	33.8	757	234	AQ0869312	AQ0869312 nbe0034P						
C 6	25.8	33.5	453	11	AA752920	AA752920 97AS0655						
C 7	25.8	33.5	905	221	CNS04SG9	AL305154 Tetraodon						
C 8	25.8	33.5	1092	222	CNS05S0L	AL351246 Tetraodon						
C 9	25.6	33.2	674	110	AW010221	AW010221 ST03E05 P						
C 10	25.4	33.0	532	106	AU001605	AU001605 AU001605						
C 11	25.2	32.7	669	245	AZ511751	AZ511751 IM0356K22						
C 12	25.2	32.7	750	164	BE187845	BE187845 CFC109-R						
C 13	25.2	32.7	897	192	AK014720	AK014720 Mus muscu						
C 14	25	32.5	852	15	AI069008	AI069008 mgae0005a						
C 15	24.8	32.2	546	258	TA144C03Q	AL466861 T. brucei						
C 16	24.8	32.2	551	258	TA94B05P	AL460030 T. brucei						
C 17	24.8	32.2	924	153	BG390246	BG390246 602416203						
C 18	24.8	32.2	1037	220	CNS01PQO	AL155152 Anopheles						
C 19	24.8	32.2	1119	172	BG023959	BG023959 602303658						
C 20	24.6	31.9	603	164	BE187906	BE187906 CFC170-R						
C 21	24.6	31.9	758	242	AZ381156	AZ381156 IM0137I04						
C 22	24.6	31.9	1027	221	CNS04EDN	AL286916 Tetraodon						
C 23	24.4	31.7	331	109	AV046350	AV046350 AV046350						
C 24	24.4	31.7	464	234	AQ050567	AQ050567 LMAJFV1.1						
C 25	24.4	31.7	553	14	AA973126	AA973126 op45f08.S						
C 26	24.4	31.7	560	222	FR0044187	AL131679 Fugu rubr						
C 27	24.4	31.7	888	175	BG282954	BG282954 602406129						
C 28	24.2	31.4	297	223	AQ070614	AQ070614 HS_2246_A						
C 29	24.2	31.4	804	239	AZ210144	AZ210144 SP_0155_A						
C 30	24.2	31.4	1942	223	AQ012198	AQ012198 463H1F053						
C 31	24	31.2	253	161	BB586972	BB586972 BB586972						
C 32	24	31.2	395	173	BG093810	BG093810 mab44e06.						
C 33	24	31.2	479	175	BG233860	BG233860 daa29e01.						
C 34	24	31.2	502	17	AL187487	AL187487 EST276 Ma						
C 35	24	31.2	509	154	BG513396	BG513396 dae04b05.						
C 36	24	31.2	513	120	AW764504	AW764504 da86a07.x						
C 37	24	31.2	538	235	AQ925334	AQ925334 RPI-23-2						
C 38	24	31.2	692	174	BG112491	BG112491 602282086						
C 39	24	31.2	731	221	CNS03QFO	AL255887 Tetraodon						
C 40	24	31.2	802	137	BE544177	BE544177 601076729						
C 41	24	31.2	906	218	AF094973	AF094973 AF094973						
C 42	24	31.2	1112	218	AF095406	AF095406 AF095406						
C 43	23.8	30.9	266	235	AQ904529	AQ904529 GSSTC0415						
C 44	23.8	30.9	492	258	TA100E09Q	AL459350 T. brucei						
C 45	23.8	30.9	543	258	TA289C09Q	AL486560 T. brucei						
ALIGNMENTS												
RESULT 1	TA382G08P/c	TA382G08P	510 bp	DNA	GSS	13-DEC-2000						
LOCUS		T. brucei sheared genomic DNA clone 382g08, forward sequence,										
DEFINITION		genomic survey sequence.										
ACCESSION	AL498114											
VERSION	AL498114.1	GI:11873836										
KEYWORDS	GSS.											
SOURCE	Trypanosoma brucei.											
ORGANISM	Trypanosoma brucei.											
REFERENCE	1	(bases 1 to 510)										
AUTHORS	Hall, N., Bowman, S., Lennard, N. J., Doggett, J., Atkin, R., Chillingworth, C., Ormond, D., Harris, B., El-Sayed, N., Hou, L., Melville, S. E., Rajandream, M. A. and Barrell, B. G.											
TITLE	Direct Submission											

JOURNAL	Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and nh@sanger.ac.uk
COMMENT	Constructed at the Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from a cloned population of Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared to give a tight size distribution (4 kb). The v + i method used for the library construction is described in detail in Smith, H. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barrell, Oxford University Press, 1999). Email: nelsayed@tigr.org Details of T. brucei sequencing at the Sanger Centre are available at http://www.sanger.ac.uk/Projects/T_brucei/. Location/Qualifiers 1..510 /organism="Trypanosoma brucei" /strain="TREU927" /db_xref="taxon:5691" /clone="382g08"
FEATURES	source 112 a 139 c 139 g 120 t
BASE COUNT	
ORIGIN	
Query Match 37.1%; Score 28.6; DB 258; Length 510; Best Local Similarity 64.2%; Pred. No. 3.4; Matches 43; Conservative 0; Mismatches 24; Indels 0; Gaps 0;	
QY 1 c t c t a a a a t g t t c a a c c g t c t c t c g c t c a c t c g g a c c g c g a a a g c g c c c 60	
Db 269 C T G A A A A A C T G C G C A A G T T G T C G C T C C G C T C G T G T C A C C G C T C C G C G A G C C C 210	
QY 61 t t a a c c a 67	
Db 209 T T C A C C A 203	
RESULT 2	
BE620662/c	610 bp mRNA EST 20-OCT-2000
LOCUS	601483044T1 NIH_MGC_69 Homo sapiens cDNA clone IMAGE:3886139 3',
DEFINITION	mRNA sequence.
ACCESSION	BE620662
VERSION	BE620662.1 GI:9891600
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	NIH-MGC http://mgc.nci.nih.gov/.
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL	Unpublished (1999)
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgapbs@mail.nih.gov Tissue Procurement: DCTD/DTP/Gazdar
COMMENT	cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at: http://image.lnl.gov Plate: L1AM9662 row: h column: 12 High quality sequence start: 22 High quality sequence stop: 610. High quality sequence stop: 610.
FEATURES	Location/Qualifiers 1..610 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:3886139" /clone_lib="NIH_MGC 69"

JOURNAL	COMMENT
Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and nhls@sanger.ac.uk	Constructed at the Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from a cloned population of Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared to give a tight size distribution (4 kb). The v + i method used for the library construction is described in detail in Smith, H. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barrell, Oxford University Press, 1999).
Details of T. brucei sequencing at the Sanger Centre are available at http://www.sanger.ac.uk/Projects/T_brucei/.	Location/Qualifiers
1. .510	/organism="Trypanosoma brucei"
	/strain="TREU927"
	/db_xref="taxon:5691"
	/clone="382g08"
BASE COUNT 112 a 139 c 139 g 120 t	
ORIGIN	
Query Match 37.1%; Score 28.6; DB 258; Length 510;	
Best Local Similarity 64.2%; Pred.No.3.4;	
Matches 43; Conservative 0; Mismatches 24; Indels 0; Gaps 0;	
QY 1 ctctacaacatggttcacacgcttccttcgctcactcgacccgacgcgacgcgcgc 60	
DB 269 CTGGAACAACTGGCGCAAGTGTTCGCTCGCTCGTGTGTCACCGCTCCGCGACGCC 210	
QY 61 ttaacca 67	
DB 209 TTCACCA 203	
RESULT 2	
BE620662/c	
LOCUS	BE620662 610 bp mRNA EST 20-OCT-2000
DEFINITION	601483644T1 NIH_MGC_69 Homo sapiens cDNA clone IMAGE:3886139 3', mRNA sequence.
ACCESSION	BE620662
VERSION	BE620662.1
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	1 (bases 1 to 610)
AUTHORS	NIH-MGC http://mgc.nci.nih.gov/.
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL	Unpublished (1999)
COMMENT	Contact: Robert Strausberg, Ph.D. Email: rcgaps@remail.nih.gov Tissue Procurement: DCTD/DPG/Gazdar cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) cDNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LLAM9662 row: h column: 12 High quality sequence start: 22 High quality sequence stop: 610.
FEATURES	source
1. .610	Location/Qualifiers
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/clone="IMAGE:3886139"
	/clone_lib="NIH_MGC_69"

```
/tissue_type="large cell carcinoma, undifferentiated"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lung; Vector: pCMV-SPORT6; Site:1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.1 kb. Library constructed by Life
Technologies."
BASE COUNT      86 a 186 c 204 g 134 t
ORIGIN

Query Match      35.8%; Score 27.6; DB 138; Length 610;
Best Local Similarity 63.6%; Pred. No. 8;
Matches 42; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

Qy 9 aatggtcaagccgttccttcgtcactcggaccgactaaagccgccccttaaccaa 68
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 84 AAGATTTCAAGCCTTCGCTAGTCTCCGTATGCGCCGCGCAACGCCCGCTCTCGGCAT 25

Qy 69 acgtta 74
||| |||
Db 24 ACGGTA 19

RESULT 3
AL524576      536 bp mRNA EST 13-FEB-2001
LOCUS AL524576 LTI_NFL003_NBC3 Homo sapiens cDNA clone CS0DC007YG04 5
DEFINITION prime, mRNA sequence.
ACCESSION AL524576
VERSION AL524576.1 GI:12788069
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 536)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES
source
1..536
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DC007YG04"
/clone_lib="LTI_NFL003_NBC3"
/sex="male"
/tissue_type="neuroblastoma cells"
/lab_host="DH10B"
/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed
by Life Technologies, Contact: Feng Liang Life
Technologies, a division of Invitrogen 9800 Medical Center
Drive Rockville, Maryland 20850, USA Fax : (1) 301 610
8371 Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com"
BASE COUNT      96 a 172 c 178 g 74 t 16 others
ORIGIN

Query Match      35.1%; Score 27; DB 105; Length 536;
Best Local Similarity 57.5%; Pred. No. 13;
Matches 42; Conservative 3; Mismatches 28; Indels 0; Gaps 0;

Qy 3 ctaacaatggttcaagccttcgttcgtcactcgtggaccggttaagccggccctt 62
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 241 CTAMCARCTGGTGACGGCGGCGCTTCAMGCTCTGGGGCCGCGAGCGCGGCGCTG 300
```

```
Qy 63 aaccaaagcttag 75
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 301 GACCAAGTCGCGAG 313

RESULT 4
AA203709      675 bp mRNA EST 24-JAN-1997
LOCUS AA203709
DEFINITION z52d10.r1 Soares_fetal_liver_spleen_lNFLS_S1 Homo sapiens cDNA
clone IMAGE:446131 5', similar to contains element OFR repetitive
element ; mRNA sequence.
ACCESSION AA203709
VERSION AA203709.1 GI:1799436
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 675)
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman
,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J.,
Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston
,R., Williamson,A., Wohldmann,P. and Wilson,R.
TITLE The WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -28M13 rev2 from Amersham
High quality sequence stop: 492.
FEATURES
source
1..675
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:446131"
/clone_lib="Soares_fetal_liver_spleen_lNFLS_S1"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)
with a modified polylinker; Site:1: Pac I; Site_2: Eco RI;
This is a subtracted version of the original Soares fetal
liver spleen lNFLS library. 1st strand cDNA was primed
with a Pac I - oligo(dT) primer [5'
AACTGGAAGAAATAATTAAGATCTTTTTTTTTTTTTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Pac I and cloned into the Pac I
and Eco RI sites of the modified pT7T3 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT      191 a 159 c 138 g 185 t 2 others
ORIGIN

Query Match      33.8%; Score 26; DB 3; Length 675;
Best Local Similarity 67.3%; Pred. No. 31;
Matches 35; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

Qy 11 tggttcaagcgttcgttcgttcactcgtggaccggttaagccggccctt 62
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 562 TGTTTCAAGCCTGTGCTCTGCTGTCACGCCGACCGTGGAGGCCNTACNTT 613

RESULT 5
AQ869312      757 bp DNA GSS 03-NOV-1999
LOCUS AQ869312
DEFINITION nbcb0034P16r CUGI Rice BAC Library (EcoRI) Oryza sativa genomic
```

```

clone nbhe0034P16r, DNA sequence.
AA0869312
DEFINITION A0869312.1 GI:6219763
ACCESSION Oryza sativa.
VERSION Oryza sativa.
KEYWORDS GSS.
SOURCE Oryza sativa.
ORGANISM Oryza sativa.
REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
AUTHORS Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
TITLE Ehrhartoidae; Oryzeae; Oryza.
JOURNAL 1 (bases 1 to 757)
COMMENT A BAC End Sequencing Framework to Sequence the Rice Genome
Contact: Wing RA
Unpublished (1998)
Clemson University Genomics Institute
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Seq primer: GGAACAGCTATGACCATG
Class: BAC ends
High quality sequence start: 25
High quality sequence stop: 448.
Location/Qualifiers
FEATURES
source
1..757
/organism="Oryza sativa"
/strain="Japonica"
/cultivar="Nipponbare"
/db_xref="taxon:4530"
/clone="nbhe0034P16r"
/clone_lib="CUGI Rice BAC Library (EcoRI)"
/tissue_type="Leaf"
/lab_host="E. coli DH10B"
/notes="Vector: pBACIndigo; Site.1: EcoRI; Site.2: EcoRI;
Rice is the most important food crop in the world. Half of
the world population, especially those inhabiting highly
populated areas of the humid tropics and subtropics, rely
on rice as their primary source of carbohydrate.
Monocotyledonous rice is a diploid plant (2n=24) with a
haploid genome equivalent of 431 Mbp (Arumuganathan and
Earle, 1991). The relatively small genome of rice, three
times larger than that of Arabidopsis, makes it suitable
for genomic studies. In order to facilitate positional
cloning, physical mapping and genome sequencing of rice,
we have constructed a BAC library from Oryza sativa.
Nipponbare variety using EcoRI as the cloning enzyme. The
library contains 55,296 clones with an average insert size
of 121 kb providing approximately 15 haploid genome
equivalents. The deep coverage allows the isolation a
particular sequence with a probability of 99.9 %. Three
high density filters, each containing 18,432 clones
(doubly spotted), represent the whole library for colony
screening and can be requested from the Clemson University
BAC/EST Resource Center (www.genome.clemson.edu)."
BAC/EST Resource Center (www.genome.clemson.edu)."
BASE COUNT 157 a 219 c 229 g 151 t 1 others
ORIGIN

Query Match 33.8%; Score 26; DB 234; Length 757;
Best Local Similarity 62.1%; Pred. No. 32;
Matches 41; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

QY 12 gggtcaagcgttcgttcgctcactcgagccggctaaagccggcccttaaccacacg 71
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 671 GGTTCGCCACTTCTTTTGGCGGACTCGGAGACGCGTCGAGCGCGCGCTGAGACCAG 730
QY 72 ttagag 77
|||||
DB 731 TTCGCG 736

RESULT 6

```

```

AA752920/c
LOCUS AA752920 453 bp mRNA EST 20-JAN-1998
DEFINITION 97AS0655 Rice Immature Seed Lambda ZAPII cDNA Library Oryza sativa
CDNA clone 97AS0655, mRNA sequence.
ACCESSION AA752920
VERSION AA752920.1 GI:2799931
KEYWORDS EST.
SOURCE Oryza sativa.
ORGANISM Oryza sativa.
REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
AUTHORS Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
TITLE Ehrhartoidae; Oryzeae; Oryza.
JOURNAL 1 (bases 1 to 453)
COMMENT Nahm,B.H., Kim,J.K., Cheong,J.J., Kim,S.I., Hahn,T.R., Moon,E.P.,
Kim,W.T., Kim,W.Y., Yang,M.S., Park,R.D., Sohn,U.I., Kang,K.Y., Lee
,M.C. and Eun,M.Y.
Large-scale Sequencing Analysis of ESTs from Rice Immature Seed
Unpublished (1998)
Contact: Eun M.Y.
Department of Cytogenetics
National Inst. of Agri. Sci. and Tech, RDA
Suwon, Kyunggido, Korea
Tel: 82 331 290 0301
Fax: 82 331 290 0307
Email: myeun@sun20.asti.re.kr
Submitted by Baek Hie Nahm, Dept of Biological Science, Myongji
University, Yongin, Korea. 449-728 bhna@bioserver.myongji.ac.kr
Seq primer: M13 Reverse Primer.
Location/Qualifiers
FEATURES
source
1..453
/organism="Oryza sativa"
/cultivar="Milyang23"
/db_xref="taxon:4530"
/clone="97AS0655"
/clone_lib="Rice Immature Seed Lambda ZAPII cDNA Library"
/tissue_type="Immature Seed"
/dev_stage="5 days after pollination"
/lab_host="E. coli SOLR"
/notes="Vector: pBluescript SK(+); Site.1: EcoRI; Site.2:
XhoI; Directional cDNA library inserted into lambda ZAPII
vector at 5' end with EcoRI and 3' end with Xho I site."
BASE COUNT 96 a 112 c 117 g 124 t
ORIGIN

Query Match 33.5%; Score 25.8; DB 11; Length 453;
Best Local Similarity 61.9%; Pred. No. 35;
Matches 39; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

QY 4 taacaaatggttcagccgttcgttcgctcactcgagccggctaaagccggccctta 63
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 422 TAAATTTTGTTCGGNGCGTTCCTTTGGGAACCTTNGACGGGAACGCCGCCCAAAA 363
QY 64 acc 66
|||
DB 362 CCC 360

RESULT 7
CNS04SG9
LOCUS CNS04SG9 905 bp DNA GSS 24-MAY-2000
DEFINITION Tetraodon nigroviridis genome survey sequence T7 end of clone
007K24 of library H from Tetraodon nigroviridis, genomic survey
sequence.
ACCESSION AL305154
VERSION AL305154.1 GI:8196052
KEYWORDS GSS; genome survey sequence.
SOURCE Tetraodon nigroviridis.
ORGANISM Tetraodon nigroviridis.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodon.

```



```

REFERENCE
AUTHORS      1 (bases 1 to 905)
              Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
              Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
              Weissenbach,J.
TITLE        Characterization and repeat analysis of the compact genome of the
              freshwater pufferfish Tetraodon nigroviridis
JOURNAL      Unpublished
AUTHORS      2 (bases 1 to 905)
              Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
              Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
              Saurin,W. and Weissenbach,J.
TITLE        Human gene number estimate provided by genome wide analysis using
              Tetraodon nigroviridis DNA sequence
JOURNAL      Unpublished
AUTHORS      3 (bases 1 to 905)
              Direct Submission
TITLE        Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
JOURNAL      This sequence is a single read and was generated as part of a large
COMMENT      scale clone-end sequencing project of the Tetraodon nigroviridis
              genome. For more information, please take a look at
              http://www.genoscope.cns.fr/tetraodon.
FEATURES     Location/Qualifiers
              source
                1..905
                /organism="Tetraodon nigroviridis"
                /db_xref="taxon:99883"
                /clone="007K24"
                /clone_lib="H"
                /note="Genoscope sequence ID : COBH007BF12XD1-end : T7"
BASE COUNT   233 a 305 c 207 g 160 t
ORIGIN
Query Match 33.5%; Score 25.8; DB 221; Length 905;
Best Local Similarity 63.9%; Pred. No. 38;
Matches 39; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

Qy 1 ctctacaaatggttcacgccttcgctcactcggtacgcgcgtacacggcgccccc 60
Db 375 ctccacaaatggttcacgccttcgctcactcggtacgcgcgtacacggcgccccc 434

Qy 61 t 61
Db 435 T 435

RESULT 8
CNS05S0L 1092 bp DNA GSS 26-MAY-2000
LOCUS Tetraodon nigroviridis genome survey sequence T3 end of clone
DEFINITION 007C15 of library A from Tetraodon nigroviridis, genomic survey
              sequence.
ACCESSION AL351246
VERSION AL351246.1 GI:8245016
KEYWORDS GSS; genome survey sequence.
SOURCE Tetraodon nigroviridis.
ORGANISM Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodon.
1 (bases 1 to 1092)
Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
Weissenbach,J.
TITLE Characterization and repeat analysis of the compact genome of the
JOURNAL freshwater pufferfish Tetraodon nigroviridis
AUTHORS 2 (bases 1 to 1092)
Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using

```

```

Tetraodon nigroviridis DNA sequence
Unpublished
REFERENCE 3 (bases 1 to 1092)
AUTHORS Direct Submission
TITLE Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
JOURNAL This sequence is a single read and was generated as part of a large
COMMENT scale clone-end sequencing project of the Tetraodon nigroviridis
              genome. For more information, please take a look at
              http://www.genoscope.cns.fr/tetraodon.
FEATURES     Location/Qualifiers
              source
                1..1092
                /organism="Tetraodon nigroviridis"
                /db_xref="taxon:99883"
                /clone="007C15"
                /clone_lib="A"
                /note="Genoscope sequence ID : COAA007AB08A1-end : T3"
BASE COUNT   263 a 367 c 240 g 207 t 15 others
ORIGIN
Query Match 33.5%; Score 25.8; DB 222; Length 1092;
Best Local Similarity 63.9%; Pred. No. 39;
Matches 39; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

Qy 1 ctctacaaatggttcacgccttcgctcactcggtacgcgcgtacacggcgccccc 60
Db 820 ctccacaaatggttcacgccttcgctcactcggtacgcgcgtacacggcgccccc 879

Qy 61 t 61
Db 880 T 880

RESULT 9
AW010221 674 bp mRNA EST 10-SEP-1999
LOCUS ST03E05 Pine Triplex shoot tip library Pinus taeda cDNA clone
DEFINITION ST03E05, mRNA sequence.
ACCESSION AW010221
VERSION AW010221.1 GI:5858999
KEYWORDS EST.
SOURCE loblolly pine.
ORGANISM Pinus taeda
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus; Pinus.
1 (bases 1 to 674)
Whetten,R.W., Kinlaw,C.S., Retzel,E. and Sederoff,R.R.
The Pine Gene Discovery Project
Unpublished (1999)
Contact: Ross Whetten
Forest Biotechnology Group
North Carolina State University
Dept. of Forestry, NC State University, 6113 Jordan Hall, Raleigh
,NC, 27695-8008
Tel: 919-515-7800
Fax: 919-515-7801
Email: rosswhetten@unity.ncsu.edu
Seq primer: 5' lambda Triplex2 Sequencing Primer.
FEATURES     Location/Qualifiers
              source
                1..674
                /organism="Pinus taeda"
                /db_xref="taxon:3352"
                /clone="ST03E05"
                /clone_lib="Pine Triplex shoot tip library"
                /note="Organ: shoot tips; Vector: Lambda Triplex; Site:1:
                Sfil (A); Site:2: Sfil (B); Shoot tips (approx. 2 cm from
                apex) were collected during the spring, frozen and used
                for mRNA isolation. The SMART-PCR method (Clontech) was
                used to prepare a library from 1 ug total RNA, using the
                Lambda Triplex vector. Plasmid subclones in pTriplex were
                recovered by cre-lox excision in E. coli strain BM25.8 and

```

```

BASE COUNT      179 a      158 c      195 g      129 t      13 others
ORIGIN

Query Match      33.2%; Score 25.6; DB 110; Length 674;
Best Local Similarity 58.1%; Pred. No. 44;
Matches 43; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

Qy 4 taacaaatggttcacgcgttcctcgtcactcactcggaccgctaaagcggccctta 63
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 447 TAGCAATTGATTAAGGCGCAAGCATCCCTTATNACAAAATGCTAGCCCTCCGCTCC 506
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 64 accaaacgttag 77
    ||||| |||||
Db 507 CCCAAGNTTAG 520

RESULT 10
AU001605      532 bp      mRNA      EST      15-JAN-1999
LOCUS
DEFINITION AU001605 Bombyx mori p50(Daizo) Bombyx mori cDNA clone fbml062f,
mRNA sequence.
ACCESSION AU001605
VERSION AU001605.1 GI:4157849
KEYWORDS EST.
SOURCE Bombyx mori
ORGANISM Bombyx mori
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia
; Bombycoidea; Bombycidae; Bombyx.
REFERENCE 1 (bases 1 to 532)
AUTHORS Mita, K., Morimyo, M., Shimada, T., Okano, K. and Maeda, S.
TITLE Establishment of cDNA database of Bombyx mori
JOURNAL Unpublished (1999)
COMMENT Contact: Mita K
Genome Research Group
National Institute of Radiological Sciences
Anagawa 4-9-1, Inage, Chiba 263-8555, Japan
Email: knita@nirs.go.jp
PROJECT = 'CREST project by JST'.
FEATURES
    source
        1..532
            /organism="Bombyx mori"
            /strain="p50(Daizo)"
            /db_xref="taxon:7091"
            /clone="fbml062f"
            /clone_lib="Bombyx mori p50(Daiz~)"
BASE COUNT      84 a      146 c      192 g      110 t

Query Match      33.08; Score 25.4; DB 106; Length 532;
Best Local Similarity 61.2%; Pred. No. 50;
Matches 41; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

Qy 9 aatggttcacgcgttcgtcgtcactcgggaccgctaaagcggcccttaacaa 68
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 14 AAGCGTCCACGAGCGCGCTGTACTCGCTTCCGGTAGACCGCGCCGAATCCAA 73
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 69 acgttag 75
    |||||
Db 74 ACGTTAG 80

RESULT 11
AZ511751
LOCUS
DEFINITION LM0356K22R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
clone UUGC1M0356K22 R, DNA sequence.
ACCESSION AZ511751
VERSION AZ511751.1 GI:10693067
KEYWORDS GSS.

house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 669)
Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,
Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly
, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A.
and Wright, D., Weiss, R.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0356 row: K column: 22
Seq primer: CACACAGGAACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 669.
Location/Qualifiers
    1..669
        /organism="Mus musculus"
        /strain="C57BL/6J"
        /db_xref="taxon:10090"
        /clone="UUGC1M0356K22"
        /clone_lib="Mouse 10kb plasmid UUGC1M library"
        /sex="Male"
        /lab_host="E. Coli strain XL10-Gold, Tl-resistant, F-"
        /note="vector: PWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptored DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of PWD42 (gil4732114|gb|AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptored mouse DNA was annealed to
adaptored vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."
BASE COUNT      167 a      158 c      113 g      231 t
ORIGIN

Query Match      32.7%; Score 25.2; DB 245; Length 669;
Best Local Similarity 60.0%; Pred. No. 61;
Matches 42; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

Qy 8 aaatggttcacgcgttcgtcgtcactcgggaccgctaaagcggcccttaacaa 67
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 43 AACTGGTGGTTACTCTCTGTGCTCTTCCCAACAGCTACAGTCCACCACACCC 102
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 68 aacgttag 77
    ||||| ||
Db 103 AACGTTTAG 112

RESULT 12
BE187845
LOCUS BE187845 750 bp mRNA EST 25-SEP-2000
```


Query Match	32.7%;	Score 25.2;	DB 192;	Length 897;	
Best Local Similarity	60.0%;	Pred. No. 63;			
Matches	42;	Conservative	0;	Mismatches	28; Indels 0; Gaps 0;
Qy	8	aaatggtcaagccgttcgctcactcgcgggaccggtctaaagccgccccttaacca	67		
Db	416	AACTGGTGTACCTCTGTTGCTCTTCCAAACAGCTACAGTCCACCAACACACCC	475		
Qy	68	aacgttagag	77		
Db	476	AACGTTTGTAG	485		
RESULT 14					
LOCUS	AI069008/c				
DEFINITION	mgae0005ab05f Magnaporthe grisea Appressorium stage cDNA Library				
ACCESSION	AI069008	852 bp	mRNA	EST	09-DEC-1999
VERSION	AI069008				
KEYWORDS	EST.				
SOURCE	Magnaporthe grisea				
ORGANISM	Magnaporthe grisea				
REFERENCE	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;				
AUTHORS	Choi,W., Fang,E., Sasinowski,H., Wing,R. and Dean,R.A.				
TITLE	Expressed sequence characterization during appressorium formation				
JOURNAL	in rice blast fungus, Magnaporthe grisea				
COMMENT	Unpublished (1998)				
	Contact: Dean,R.A.				
	Clemson University Genomics Institute				
	100 Jordan Hall, Clemson University, Clemson, SC 29634, USA				
	Tel: 864 656 5737				
	Fax: 864 656 4293				
	Email: rdean@clemson.edu				
	Seq primer: T3 primer (AATTAACTCTCACTAAAGG)				
	High quality sequence stop: 386.				
FEATURES	Location/Qualifiers				
source	1..852				
	/organism="Magnaporthe grisea"				
	/strain="70-15"				
	/db_xref="taxon:148305"				
	/clone="mgae0005ab05f"				
	/clone_lib="Magnaporthe grisea Appressorium stage cDNA				
	Library"				
	/dev_stage="Germinated conidia on appressorium-inductive				
	surface"				
	/note="Vector: pBlueScriptII SK(+) Vector; Site_1: EcoRI;				
	Site_2: XhoI; The appressorium formation-specific cDNA				
	library was constructed from conidia germinated for 5-8				
	hr on an inductive surface. The library has an average				
	insert size of 1.5 kbp."				
BASE COUNT	213 a	215 c	221 g	198 t	5 others
ORIGIN					
Query Match	32.5%;	Score 25;	DB 15;	Length 852;	
Best Local Similarity	64.9%;	Pred. No. 74;			
Matches	37;	Conservative	0;	Mismatches	20; Indels 0; Gaps 0;
Qy	5	aacaaatggttcacgcccgttcgctcactcgcgggaccggtctaaagccggcccct	61		
Db	427	AACCACTGGGTCAAACAGCTCTCTCGAACAGTCGGGAACGGCAAAAAGTGAACACT	371		
RESULT 15					
LOCUS	TA144C03Q	546 bp	DNA	GSS	13-DEC-2000
DEFINITION	T. brucei sheared genomic DNA clone 144c03, reverse sequence,				
	genomic survey sequence.				
ACCESSION	AL466861				
VERSION	AL466861.1	GI:11836216			
KEYWORDS	GSS.				
SOURCE	Trypanosoma brucei.				
ORGANISM	Trypanosoma brucei				
REFERENCE	Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;				
AUTHORS	1 (bases 1 to 546)				
TITLE	Hall,N., Bowman,S., Lennard,N.J., Doggett,J., Atkin,R.,				
JOURNAL	Chillingworth,C., Ormond,D., Harris,B., El-Sayed,N., Hou,L.,				
	Melville,S.E., Rajandream,M.A. and Barrell,B.G.				
	Direct Submission				
	Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing				
	project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton,				
	Cambridge CB10 1SA, E-mail: barrellesanger.ac.uk and				
	nhlsanger.ac.uk				
	Constructed at the Institute for Genomic Research (TIGR),				
	Rockville, MD. Genomic DNA isolated from a cloned population of				
	Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared				
	to give a tight size distribution (
	4 kb). The v + i method used for the library construction is				
	described in detail in Smith, H. and Venter, J.C. (Making small				
	insert libraries for whole genome shotgun sequencing projects. In				
	Genome Sequencing: A Practical Approach, eds. M. Vaudin and B.				
	Barrell, Oxford University Press, 1999).				
	Email: nelsayed@tigr.org				
	Details of T. brucei sequencing at the Sanger Centre are available				
	at http://www.sanger.ac.uk/Projects/T_brucei/.				
FEATURES	Location/Qualifiers				
source	1..546				
	/organism="Trypanosoma brucei"				
	/strain="TREU927"				
	/db_xref="taxon:5691"				
	/clone="144c03"				
BASE COUNT	131 a	162 c	112 g	141 t	
ORIGIN					
Query Match	32.2%;	Score 24.8;	DB 258;	Length 546;	
Best Local Similarity	60.3%;	Pred. No. 83;			
Matches	41;	Conservative	0;	Mismatches	27; Indels 0; Gaps 0;
Qy	4	taacaaatggttcacgcccgttcgctcactcgcgggaccggtctaaagccggcccctta	63		
Db	31	TAATAATCATTCACACTCATCCGACCACTTAGTCGGCATTTGGCTGCAGCGAGCTCTCA	90		
Qy	64	accaaaacg	71		
Db	91	TCCATGCG	98		
Search completed: August 2, 2001, 18:38:56					
Job time: 6346 sec					

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 2, 2001, 16:53:11 ; Search time 49.77 Seconds
(without alignments)
1081.043 Million cell updates/sec

Title: US-09-701-626A-78

Perfect score: 77
Sequence: 1 cctcaacaatggtcaagc.....cccttaacacaaacgtagtag 77

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 327855 seqs, 349373361 residues

Total number of hits satisfying chosen parameters: 655710

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summa: ies

Database : Pending Patents_NA_New**

- 1: /cgnl_7/ptodata/2/pna/PCT_NEW_COMB.seq:**
- 2: /cgnl_7/ptodata/2/pna/US06_NEW_COMB.seq:**
- 3: /cgnl_7/ptodata/2/pna/US07_NEW_COMB.seq:**
- 4: /cgnl_7/ptodata/2/pna/US08_NEW_COMB.seq:**
- 5: /cgnl_7/ptodata/2/pna/US09_NEW_COMB.seq:**
- 6: /cgnl_7/ptodata/2/pna/US60_NEW_COMB.seq:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	26.4	34.3	474	5	US-09-758-472-1145
C 2	25.2	32.7	12595	5	US-09-764-902-2226
C 3	25	32.5	514	5	US-09-764-864-389
C 4	25	32.5	629	5	US-09-764-864-797
C 5	24.8	32.2	32193	5	US-09-760-471-110
C 6	24.8	32.2	32193	5	US-09-760-485-1268
C 7	24.8	32.2	32194	5	US-09-760-471-109
C 8	24.8	32.2	32194	5	US-09-760-485-1267
C 9	23.8	30.9	7727	5	US-09-764-891-7685
C 10	23.8	30.9	125803	5	US-09-803-736-621
C 11	23.4	30.4	1449	5	US-09-803-110-3192
C 12	23.4	30.4	184668	5	US-09-803-110-207
C 13	22.6	29.4	654	6	US-60-278-258-7789
C 14	22.6	29.4	2992	5	US-09-898-888-5715
C 15	22.6	29.4	55739	5	US-09-803-736-485
C 16	22.6	29.4	58690	5	US-09-803-736-209
C 17	22.6	29.4	81414	5	US-09-803-736-757
C 18	22.6	29.4	95214	5	US-09-803-736-382
C 19	22.6	29.4	110157	5	US-09-803-736-758
C 20	22.4	29.1	312	5	US-09-803-110-7068
C 21	22.4	29.1	876	5	US-09-803-110-3052
C 22	22.4	29.1	1149	5	US-09-758-439-253
C 23	22.4	29.1	3116	5	US-09-764-872-585
C 24	22.4	29.1	37604	5	US-09-803-110-194
C 25	22.4	29.1	542340	5	US-09-803-110-219
C 26	22.2	28.8	1447	5	US-09-760-475-928
C 27	22.2	28.8	1447	5	US-09-760-485-302

28	22.2	28.8	2192	6	US-60-278-258-14693	Sequence 14693, A
29	22.2	28.8	3788	5	US-09-347-316-33	Sequence 33, Appl
30	22.2	28.8	12269	5	US-09-760-475-3882	Sequence 3882, Ap
31	22.2	28.8	12269	5	US-09-760-485-1308	Sequence 1308, Ap
32	22.2	28.8	12548	5	US-09-760-475-3881	Sequence 3881, Ap
C 33	22.2	28.8	90632	5	US-09-803-736-1160	Sequence 1160, Ap
C 34	22	28.6	1510	5	US-09-764-891-7302	Sequence 7302, Ap
C 35	22	28.6	1510	5	US-09-764-891-7303	Sequence 7303, Ap
36	22	28.6	1510	5	US-09-764-891-8373	Sequence 8373, Ap
37	22	28.6	1510	5	US-09-764-891-8374	Sequence 8374, Ap
38	22	28.6	32222	5	US-09-764-887-619	Sequence 619, App
C 39	21.8	28.3	345	5	US-09-804-730-15786	Sequence 15786, A
C 40	21.8	28.3	675	5	US-09-758-474-612	Sequence 612, App
C 41	21.8	28.3	104364	5	US-09-803-736-337	Sequence 337, App
42	21.8	28.3	318095	5	US-09-803-110-215	Sequence 215, App
C 43	21.6	28.1	142	5	US-09-804-730-8268	Sequence 8268, Ap
C 44	21.6	28.1	474	6	US-60-255-619-14683	Sequence 14683, A
45	21.6	28.1	626	5	US-09-757-031-90	Sequence 90, Appl

ALIGNMENTS

RESULT 1
US-09-758-472-1145/c
; Sequence 1145, Application US/09758472
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PH001
; CURRENT APPLICATION NUMBER: US/09/758,472
; PRIOR FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: 60/179,065
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/180,628
; PRIOR FILING DATE: 2000-02-04
; NUMBER OF SEQ ID NOS: 9632
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1145
; LENGTH: 474
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (382)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (442)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (453)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-758-472-1145

Query Match 34.3%; Score 26.4; DB 5; Length 474;
Best Local Similarity 61.8%; Pred. No. 0.67;
Matches 42; Conservative 0; Mismatches 26; Indels 0; Gaps 0;
Qy 10 atggttaagccgttccttcgtcactcgggacccggtacaaagccgccccttaaccacaa 69
|| ||||| ||||| | ||||| || ||||| ||||| |||||
Db 131 ATAGTCTTACCGTTCGTGTAATTTCTCGTGAACTTTGATCCGCCCTCCGGCAA 72
Qy 70 cgttagag 77
| ||||
Db 71 TATAAGAG 64

RESULT 2
US-09-764-902-2226/c
; Sequence 2226, Application US/09764902
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.


```
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1288
; LENGTH: 32193
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-760-485-1268

Query Match      32.2%; Score 24.8; DB 5; Length 32193;
Best Local Similarity 67.3%; Pred. No. 7.7;
Matches 35; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY 24 tcgttcgtcactcggagccggttaagccggccccccttaacaaacgtag 75
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6198 TCTCCTCCCTCCCGCGGACTGGGTAAAGCCAGCCCTGTAGTGAATGTAG 6147

RESULT 7
US-09-760-471-109/c
; Sequence 109, Application US/09760471
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT240
; CURRENT APPLICATION NUMBER: US/09/760,471
; CURRENT FILING DATE: 2001-01-16
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 131
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 109
; LENGTH: 32194
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-760-471-109

Query Match      32.2%; Score 24.8; DB 5; Length 32194;
Best Local Similarity 67.3%; Pred. No. 7.7;
Matches 35; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY 24 tcgttcgtcactcggagccggttaagccggccccccttaacaaacgtag 75
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 7278 TCTCCTCCCTCCCGCGGACTGGGTAAAGCCAGCCCTGTAGTGAATGTAG 7227

RESULT 8
US-09-760-485-1267/c
; Sequence 1267, Application US/09760485
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PJ217
; CURRENT APPLICATION NUMBER: US/09/760,485
; CURRENT FILING DATE: 2001-01-16
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1477
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1267
; LENGTH: 32194
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-760-485-1267

Query Match      32.2%; Score 24.8; DB 5; Length 32194;
Best Local Similarity 67.3%; Pred. No. 7.7;
Matches 35; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY 24 tcgttcgtcactcggagccggttaagccggccccccttaacaaacgtag 75
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 7278 TCTCCTCCCTCCCGCGGACTGGGTAAAGCCAGCCCTGTAGTGAATGTAG 7227
```

```
RESULT 9
US-09-764-891-7685/c
; Sequence 7685, Application US/09764891
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7685
; LENGTH: 7727
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-891-7685

Query Match      30.9%; Score 23.8; DB 5; Length 7727;
Best Local Similarity 57.3%; Pred. No. 13;
Matches 43; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 2 tctaacaaatggttcgaagccggttcgttcgtcactcgggaccggttaagccgccccct 61
|| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3427 TCATACAAATCAGTCCTTTCAAATGCTTCTGTCACCTGGGCTGAGTCCAAAGGACACCTT 3468

QY 62 taacaaacggttaga 76
| | ||| ||| |||
Db 3367 TCCCAAAACATTACA 3353

RESULT 10
US-09-803-736-621
; Sequence 621, Application US/09803736
; GENERAL INFORMATION:
; APPLICANT: Bush, David F.
; APPLICANT: Levin, Irena M.
; APPLICANT: Norris, Susan R.
; APPLICANT: Rounsley, Steven D.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Plant Polymorphic Markers and Uses Thereof
; FILE REFERENCE: 38-10(15493)D
; CURRENT APPLICATION NUMBER: US/09/803,736
; CURRENT FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 09/534,859
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: identified by Attorney Docket number 04983.0206CPUS01 38
; PRIOR FILING DATE: 2000-10-20
; NUMBER OF SEQ ID NOS: 1582
; SEQ ID NO 621
; LENGTH: 125803
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-803-736-621

Query Match      30.9%; Score 23.8; DB 5; Length 125803;
Best Local Similarity 62.7%; Pred. No. 26;
Matches 37; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

QY 6 acaaatggttcgaagccggttcgttcgtcactcgggaccggttaagccgccccctaa 64
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 52099 acaattgatcagggtcgtctctatgctctctcccccacatgatgactcgggtccatta 52157

RESULT 11
US-09-803-110-3192
; Sequence 3192, Application US/09803110
; GENERAL INFORMATION:
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; TITLE OF INVENTION: Agrobacterium tumefaciens Genome Sequences and Uses Thereof
```



```
; FILE REFERENCE: 38-10(15490)D
; CURRENT APPLICATION NUMBER: US/09/803,110
; CURRENT FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 09/739,449
; PRIOR FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: US 09/514,000
; PRIOR FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: US 60/168,139
; PRIOR FILING DATE: 1999-12-01
; NUMBER OF SEQ ID NOS: 13351
; SEQ ID NO 3192
; LENGTH: 1449
; TYPE: DNA
; ORGANISM: Agrobacterium tumefaciens
US-09-803-110-3192

Query Match      30.4%; Score 23.4; DB 5; Length 1449;
Best Local Similarity 60.0%; Pred. No. 12;
Matches 39; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 9 aatggttcaagccgttcgctcactcgtggaacggcgttaagcggcccttaaccaa 68
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 391 aatgtctataccggctccggttcctccgcccccggtcgtggtgaagcgggccacttggccag 450

QY 69 acgtt 73
||| |

Db 451 ccgat 455

RESULT 12
US-09-803-110-207
; Sequence 207, Application US/09803110
; GENERAL INFORMATION:
; APPLICANT: Hinkie, Gregory J.
; TITLE OF INVENTION: Agrobacterium tumefaciens Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15490)D
; CURRENT APPLICATION NUMBER: US/09/803,110
; CURRENT FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 09/739,449
; PRIOR FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: US 09/514,000
; PRIOR FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: US 60/168,139
; PRIOR FILING DATE: 1999-12-01
; NUMBER OF SEQ ID NOS: 13351
; SEQ ID NO 207
; LENGTH: 184668
; TYPE: DNA
; ORGANISM: Agrobacterium tumefaciens
US-09-803-110-207

Query Match      30.4%; Score 23.4; DB 5; Length 184668;
Best Local Similarity 60.0%; Pred. No. 40;
Matches 39; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 9 aatggttcaagccgttcgctcactcgtggaacggcgttaagcggcccttaaccaa 68
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 161493 aatgtctataccggctccggttcctccgcccccggtcgtggtgaagcgggccacttggccag 161552

QY 69 acgtt 73
||| |

Db 161553 ccgat 161557

RESULT 13
US-60-278-7789/c
; Sequence 7789, Application US/60278258
; GENERAL INFORMATION:
; APPLICANT: Morris, MacDonald
; APPLICANT: Lal, Preeti
```

```
; APPLICANT: Diep, Dinh
; TITLE OF INVENTION: Method for the Identification of Sequence Polymorphisms Using
; TITLE OF INVENTION: Polynucleotide Sequence Databases, and Single Nucleotide
; TITLE OF INVENTION: Polymorphisms Identified Thereby
; FILE REFERENCE: GX-0010-1 P
; CURRENT APPLICATION NUMBER: US/60/278,258
; CURRENT FILING DATE: 2001-03-23
; NUMBER OF SEQ ID NOS: 17730
; SOFTWARE: PERL Program
; SEQ ID NO 7789
; LENGTH: 654
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No: 347789.26
; NAME/KEY: unsure
; LOCATION: 179
; OTHER INFORMATION: a, t, c, g, or other
US-60-278-258-7789

Query Match      29.4%; Score 22.6; DB 6; Length 654;
Best Local Similarity 63.0%; Pred. No. 21;
Matches 34; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

QY 10 atggttcaagccgttcgctcactcgtggaacggcgttaagcggccctta 63
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 227 AGGGTCCAGCTGGCTCAGCTCTCTCAGCTCAGCTACAGCCAGCCNCCTTA 174

RESULT 14
US-09-898-888-5715/c
; Sequence 5715, Application US/09898888
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL CONFIGS OBTAINED FROM VARIOUS cDNA
; FILE REFERENCE: 20411-748CON1
; CURRENT APPLICATION NUMBER: US/09/898,888
; CURRENT FILING DATE: 2001-07-03
; PRIOR APPLICATION NUMBER: 09/340,623
; PRIOR FILING DATE: 1999-06-28
; NUMBER OF SEQ ID NOS: 45207
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5715
; LENGTH: 2992
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-898-888-5715

Query Match      29.4%; Score 22.6; DB 5; Length 2992;
Best Local Similarity 55.8%; Pred. No. 30;
Matches 43; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

QY 1 ctctaaacaatggttcaagccgttcgctcactcgtggaacggcgttaagcggccccc 60
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1886 CTCACACACGATGCTCCCTGCTGACACGCTTGACATGTCATACAGGGTGAGGGGGCCAC 1827

QY 61 ttaaccaaaccgttagag 77
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1826 TGCAGCAGAGTCAGG 1810

RESULT 15
US-09-803-736-485
; Sequence 485, Application US/09803736
; GENERAL INFORMATION:
; APPLICANT: Bush, David F.
; APPLICANT: Levin, Irena M.
; APPLICANT: Norri, Susan R.
; APPLICANT: Rounsley, Steven D.
```

```

: APPLICANT: Wiegand, Roger C.
: TITLE OF INVENTION: Plant Polymorphic Markers and Uses Thereof
: FILE REFERENCE: 38-10(15493)D
: CURRENT APPLICATION NUMBER: US/09/803,736
: CURRENT FILING DATE: 2001-03-12
: PRIOR APPLICATION NUMBER: US 09/534,859
: PRIOR FILING DATE: 2000-03-29
: PRIOR APPLICATION NUMBER: identified by Attorney Docket number 04983.0206CPS01
: PRIOR FILING DATE: 2000-10-20
: NUMBER OF SEQ ID NOS: 1582
: SEQ ID NO 485
: LENGTH: 55739
: TYPE: DNA
: ORGANISM: Arabidopsis thaliana
US-09-803-736--485

```

	Query Match	29.4%	Score 22.6;	DB 5;	Length 55739;
	Best Local Similarity	58.0%;	Pred. No. 61;		
	Matches	40;	Conservative	0;	Mismatches
			29;	Indels	0;
				Gaps	0;
QY	5	acacaaatggttc	aagcgg	tgcggttc	caactc
		acatgaagttca	atcgcg	attact	gacta
Db	52119	acatgaagttca	atcgcg	attact	gacta
		acacaaatggttc	aagcgg	tgcggttc	caactc
		acacaaatggttc	aagcgg	tgcggttc	caactc
QY	65	ccaaacggtt	73		
Db	52179	ccagcggtt	52187		

Search completed: August 2, 2001, 16:54:36
Job time: 85 sec

Mon Aug 6 10:48:50 2001

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using s.w. model

Run on: August 2, 2001, 16:53:10 : Search time 2788.81 Seconds
(without alignments)
407.993 Million cell updates/sec

Title: US-09-701-626A-78
Perfect score: 77
Sequence: 1 ccttaacaaagtgttaagc.....cccttaacaaacgttagag 77

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
Searched: 14155048 seqs, 7388405095 residues 28310096
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_NA_Main:*

1: /cgnl_7/ptodata/1/pna/US06_COMB.seq:*

2: /cgnl_7/ptodata/1/pna/US06_COMB.seq:*

3: /cgnl_7/ptodata/1/pna/US07_COMB.seq:*

4: /cgnl_7/ptodata/1/pna/US08_COMB.seq:*

5: /cgnl_7/ptodata/1/pna/US081_COMB.seq:*

6: /cgnl_7/ptodata/1/pna/US082_COMB.seq:*

7: /cgnl_7/ptodata/1/pna/US083_COMB.seq:*

8: /cgnl_7/ptodata/1/pna/US084_COMB.seq:*

9: /cgnl_7/ptodata/1/pna/US085_COMB.seq:*

10: /cgnl_7/ptodata/1/pna/US086_COMB.seq:*

11: /cgnl_7/ptodata/1/pna/US087_COMB.seq:*

12: /cgnl_7/ptodata/1/pna/US088_COMB.seq:*

13: /cgnl_7/ptodata/1/pna/US089_COMB.seq:*

14: /cgnl_7/ptodata/1/pna/US090_COMB.seq:*

15: /cgnl_7/ptodata/1/pna/US091_COMB.seq:*

16: /cgnl_7/ptodata/1/pna/US092_COMB.seq:*

17: /cgnl_7/ptodata/1/pna/US093_COMB.seq:*

18: /cgnl_7/ptodata/1/pna/US094_COMB.seq:*

19: /cgnl_7/ptodata/1/pna/US095A_COMB.seq:*

20: /cgnl_7/ptodata/1/pna/US095B_COMB.seq:*

21: /cgnl_7/ptodata/1/pna/US095C_COMB.seq:*

22: /cgnl_7/ptodata/1/pna/US095D_COMB.seq:*

23: /cgnl_7/ptodata/1/pna/US096A_COMB.seq:*

24: /cgnl_7/ptodata/1/pna/US096B_COMB.seq:*

25: /cgnl_7/ptodata/1/pna/US096C_COMB.seq:*

26: /cgnl_7/ptodata/1/pna/US096D_COMB.seq:*

27: /cgnl_7/ptodata/1/pna/US096E_COMB.seq:*

28: /cgnl_7/ptodata/1/pna/US097A_COMB.seq:*

29: /cgnl_7/ptodata/1/pna/US097B_COMB.seq:*

30: /cgnl_7/ptodata/1/pna/US097C_COMB.seq:*

31: /cgnl_7/ptodata/1/pna/US098_COMB.seq:*

32: /cgnl_7/ptodata/1/pna/US098_COMB.seq:*

33: /cgnl_7/ptodata/1/pna/US6001_COMB.seq:*

34: /cgnl_7/ptodata/1/pna/US6002_COMB.seq:*

35: /cgnl_7/ptodata/1/pna/US6003_COMB.seq:*

36: /cgnl_7/ptodata/1/pna/US6004_COMB.seq:*

37: /cgnl_7/ptodata/1/pna/US6005_COMB.seq:*

38: /cgnl_7/ptodata/1/pna/US6006_COMB.seq:*

39: /cgnl_7/ptodata/1/pna/US6007_COMB.seq:*

40: /cgnl_7/ptodata/1/pna/US6008_COMB.seq:*

41: /cgnl_7/ptodata/1/pna/US6009_COMB.seq:*

42: /cgnl_7/ptodata/1/pna/US6010_COMB.seq:*

43: /cgnl_7/ptodata/1/pna/US6011_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	ID	Description
1	77	100.0	77	1	PCT-US99-13295-78
2	77	100.0	77	28	US-09-701-626A-78
3	74	96.1	77	1	PCT-US99-13295-77
4	74	96.1	77	28	US-09-701-626A-77
5	74	96.1	14143	1	PCT-US99-13295-1
6	74	96.1	14143	1	PCT-US99-13295-2
7	74	96.1	14143	28	US-09-701-626A-1
8	74	96.1	14143	1	PCT-US99-13295-4
9	71.8	93.2	77	1	PCT-US99-13295-19
10	71.8	93.2	77	28	US-09-701-626A-19
11	71.4	92.7	77	1	PCT-US99-13295-63
12	71.4	92.7	77	1	PCT-US99-13295-119
13	71.4	92.7	77	28	US-09-701-626A-63
14	71.4	92.7	77	28	US-09-701-626A-119
15	71.2	92.5	77	1	PCT-US99-13295-65
16	71.2	92.5	77	28	US-09-701-626A-65
17	69.8	90.6	77	1	PCT-US99-13295-11
18	69.8	90.6	77	28	US-09-701-626A-11
19	69.8	90.6	77	28	US-09-701-626A-24
20	69.8	90.6	77	28	US-09-701-626A-55
21	69.2	89.9	77	1	PCT-US99-13295-22
22	69.2	89.9	77	28	US-09-701-626A-22
23	68.6	89.1	77	1	PCT-US99-13295-41
24	68.6	89.1	77	28	US-09-701-626A-41
25	68.2	88.6	77	1	PCT-US99-13295-68
26	68.2	88.6	77	28	US-09-701-626A-68
27	68.2	88.6	77	28	US-09-701-626A-4
28	67.6	87.8	77	1	PCT-US99-13295-5
29	67.6	87.8	77	28	US-09-701-626A-5
30	67.4	87.5	77	1	PCT-US99-13295-3
31	67.4	87.5	77	28	US-09-701-626A-3
32	67.4	87.5	7300	28	US-09-701-626A-2
33	67.4	87.5	7300	28	US-09-701-626A-122
34	67.4	87.5	7300	28	US-09-701-626A-98
35	67.4	87.5	7300	28	US-09-701-626A-75
36	67.4	87.5	7300	28	US-09-701-626A-26
37	67.4	87.5	7300	28	US-09-701-626A-122
38	67.4	87.5	7300	28	US-09-701-626A-21
39	67.4	87.5	7300	28	US-09-701-626A-98
40	67.4	87.5	7300	28	US-09-701-626A-75
41	67.4	87.5	7300	28	US-09-701-626A-26
42	67.4	87.5	7300	28	US-09-701-626A-122
43	67.4	87.5	7300	28	US-09-701-626A-21
44	67.4	87.5	7300	28	US-09-701-626A-98
45	67.4	87.5	7300	28	US-09-701-626A-75

ALIGNMENTS

```

;RESULT 1
;CT-US99-13295-78
; Sequence 78, Application PC/TUS9913295
; GENERAL INFORMATION:
; APPLICANT: New England Biolabs, Inc.
; APPLICANT: Vaisvila, Romualdas
; APPLICANT: Morgan, Richard D.
; APPLICANT: Raleigh, Elisabeth
; TITLE OF INVENTION: Restriction Enzyme Gene Discovery Method
; FILE REFERENCE: Gene Discovery Method
; CURRENT APPLICATION NUMBER: PCT/US99/13295
; EARLIER FILING DATE: 1999-06-11
; EARLIER APPLICATION NUMBER: 60/089,086
; EARLIER FILING DATE: 1998-06-12
; EARLIER APPLICATION NUMBER: 60/089,101
; EARLIER FILING DATE: 1998-06-12
; EARLIER APPLICATION NUMBER: 60/089,101
; NUMBER OF SEQ ID NOS: 130
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 78
; LENGTH: 77
; TYPE: DNA
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: Genomic DNA of
; OTHER INFORMATION: Pseudomonas Alcaligenes NEB #585 (ATCC 55044)
; PCT-US99-13295-78

Query Match 100.0%; Score 77; DB 1; Length 77;
Best Local Similarity 100.0%; Pred. No. 4.8e-19;
Matches 77; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ccttaacaaatggttcaagcgttcgtctcactcgggacggcgtaaaagcggccccc 60
Db 1 ccttaacaaatggttcaagcgttcgtctcactcgggacggcgtaaaagcggccccc 60

Qy 61 ttaaccaaacttagag 77
Db 61 ttaaccaaacttagag 77

;RESULT 2
;US-09-701-626A-78
; Sequence 78, Application US/09701626A
; GENERAL INFORMATION:
; APPLICANT: Vaisvila, Romualdas
; APPLICANT: Morgan, Richard D.
; APPLICANT: Raleigh, Elisabeth
; TITLE OF INVENTION: Restriction Enzyme Gene Discovery Method
; FILE REFERENCE: NEB-165PUS
; CURRENT APPLICATION NUMBER: US/09/701,626A
; CURRENT FILING DATE: 2001-04-16
; PRIOR FILING DATE: 1998-06-12
; PRIOR APPLICATION NUMBER: 60/089,086
; PRIOR FILING DATE: 1998-06-12
; PRIOR APPLICATION NUMBER: 60/089,101
; PRIOR FILING DATE: 1998-06-12
; PRIOR APPLICATION NUMBER: PCT/US99/13295
; PRIOR FILING DATE: 1999-06-11
; NUMBER OF SEQ ID NOS: 130
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 78
; LENGTH: 77
; TYPE: DNA
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: Genomic DNA of
; OTHER INFORMATION: Pseudomonas Alcaligenes NEB #585 (ATCC 55044)
; US-09-701-626A-78

Query Match 100.0%; Score 77; DB 1; Length 77;
Best Local Similarity 100.0%; Pred. No. 4.8e-19;
Matches 77; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ccttaacaaatggttcaagcgttcgtctcactcgggacggcgtaaaagcggccccc 60
Db 1 ccttaacaaatggttcaagcgttcgtctcactcgggacggcgtaaaagcggccccc 60

Qy 61 ttaaccaaacttagag 77
Db 61 ttaaccaaacttagag 77

;RESULT 3
;PCT-US99-13295-77
; Sequence 77, Application PC/TUS9913295
; GENERAL INFORMATION:
; APPLICANT: New England Biolabs, Inc.
; APPLICANT: Vaisvila, Romualdas
; APPLICANT: Morgan, Richard D.
; APPLICANT: Raleigh, Elisabeth
; TITLE OF INVENTION: Restriction Enzyme Gene Discovery Method
; FILE REFERENCE: Gene Discovery Method
; CURRENT APPLICATION NUMBER: PCT/US99/13295
; EARLIER FILING DATE: 1999-06-11
; EARLIER APPLICATION NUMBER: 60/089,086
; EARLIER FILING DATE: 1998-06-12
; EARLIER APPLICATION NUMBER: 60/089,101
; EARLIER FILING DATE: 1998-06-12
; NUMBER OF SEQ ID NOS: 130
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 77
; LENGTH: 77
; TYPE: DNA
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: Genomic DNA of
; OTHER INFORMATION: Pseudomonas Alcaligenes NEB #585 (ATCC 55044)
; PCT-US99-13295-77

Query Match 96.1%; Score 74; DB 1; Length 77;
Best Local Similarity 100.0%; Pred. No. 6.7e-18;
Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 ctaacaaatggttcaagcgttcgtctcactcgggacggcgtaaaagcggccctt 62
Db 3 ctaacaaatggttcaagcgttcgtctcactcgggacggcgtaaaagcggccctt 62

Qy 63 aaccaaacttaga 76
Db 63 aaccaaacttaga 76

;RESULT 4
;US-09-701-626A-77
; Sequence 77, Application US/09701626A
; GENERAL INFORMATION:
; APPLICANT: Vaisvila, Romualdas
; APPLICANT: Morgan, Richard D.
; APPLICANT: Raleigh, Elisabeth
; TITLE OF INVENTION: Restriction Enzyme Gene Discovery Method
; FILE REFERENCE: NEB-165PUS
; CURRENT APPLICATION NUMBER: US/09/701,626A
; CURRENT FILING DATE: 2001-04-16
; PRIOR FILING DATE: 1998-06-12
; PRIOR APPLICATION NUMBER: 60/089,086
; PRIOR FILING DATE: 1998-06-12
; PRIOR APPLICATION NUMBER: 60/089,101
; PRIOR FILING DATE: 1998-06-12
; PRIOR APPLICATION NUMBER: PCT/US99/13295
; PRIOR FILING DATE: 1999-06-11
; NUMBER OF SEQ ID NOS: 130
; US-09-701-626A-78
```

```

Query Match 100.0%; Score 77; DB 28; Length 77;
Best Local Similarity 100.0%; Pred. No. 4.8e-19;
Matches 77; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ccttaacaaatggttcaagcgttcgtctcactcgggacggcgtaaaagcggccccc 60
Db 1 ccttaacaaatggttcaagcgttcgtctcactcgggacggcgtaaaagcggccccc 60

Qy 61 ttaaccaaacttagag 77
Db 61 ttaaccaaacttagag 77

;RESULT 3
;PCT-US99-13295-77
; Sequence 77, Application PC/TUS9913295
; GENERAL INFORMATION:
; APPLICANT: New England Biolabs, Inc.
; APPLICANT: Vaisvila, Romualdas
; APPLICANT: Morgan, Richard D.
; APPLICANT: Raleigh, Elisabeth
; TITLE OF INVENTION: Restriction Enzyme Gene Discovery Method
; FILE REFERENCE: Gene Discovery Method
; CURRENT APPLICATION NUMBER: PCT/US99/13295
; EARLIER FILING DATE: 1999-06-11
; EARLIER APPLICATION NUMBER: 60/089,086
; EARLIER FILING DATE: 1998-06-12
; EARLIER APPLICATION NUMBER: 60/089,101
; EARLIER FILING DATE: 1998-06-12
; NUMBER OF SEQ ID NOS: 130
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 77
; LENGTH: 77
; TYPE: DNA
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: Genomic DNA of
; OTHER INFORMATION: Pseudomonas Alcaligenes NEB #585 (ATCC 55044)
; PCT-US99-13295-77

Query Match 96.1%; Score 74; DB 1; Length 77;
Best Local Similarity 100.0%; Pred. No. 6.7e-18;
Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 ctaacaaatggttcaagcgttcgtctcactcgggacggcgtaaaagcggccctt 62
Db 3 ctaacaaatggttcaagcgttcgtctcactcgggacggcgtaaaagcggccctt 62

Qy 63 aaccaaacttaga 76
Db 63 aaccaaacttaga 76

;RESULT 4
;US-09-701-626A-77
; Sequence 77, Application US/09701626A
; GENERAL INFORMATION:
; APPLICANT: Vaisvila, Romualdas
; APPLICANT: Morgan, Richard D.
; APPLICANT: Raleigh, Elisabeth
; TITLE OF INVENTION: Restriction Enzyme Gene Discovery Method
; FILE REFERENCE: NEB-165PUS
; CURRENT APPLICATION NUMBER: US/09/701,626A
; CURRENT FILING DATE: 2001-04-16
; PRIOR FILING DATE: 1998-06-12
; PRIOR APPLICATION NUMBER: 60/089,086
; PRIOR FILING DATE: 1998-06-12
; PRIOR APPLICATION NUMBER: 60/089,101
; PRIOR FILING DATE: 1998-06-12
; PRIOR APPLICATION NUMBER: PCT/US99/13295
; PRIOR FILING DATE: 1999-06-11
; NUMBER OF SEQ ID NOS: 130
; US-09-701-626A-77
```



```

RESULT 8
PCT-US99-13295-4
; Sequence 4, Application PC/TUS9913295
; GENERAL INFORMATION:
; APPLICANT: New England Biolabs, Inc.
; APPLICANT: Vaisvila, Romualdas
; APPLICANT: Morgan, Richard D.
; APPLICANT: Raleigh, Elisabeth
; TITLE OF INVENTION: Restriction Enzyme Gene Discovery Method
; FILE REFERENCE: Gene Discovery Method
; CURRENT APPLICATION NUMBER: PCT/US99/13295
; CURRENT FILING DATE: 1999-06-11
; EARLIER APPLICATION NUMBER: 60/089,086
; EARLIER FILING DATE: 1998-06-12
; EARLIER APPLICATION NUMBER: 60/089,101
; EARLIER FILING DATE: 1998-06-12
; NUMBER OF SEQ ID NOS: 130
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 42143
; TYPE: DNA
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: Genomic DNA of
; OTHER INFORMATION: Pseudomonas Alcaligenes NEB #585 (ATCC 55044)
PCT-US99-13295-4

Query Match          96.1%; Score 74; DB 1; Length 42143;
Best Local Similarity 100.0%; Pred. No. 2.6e-17;
Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 ctactaaatggttcaaacgcttcgcttcgctcactcgggaccggtctaaagccgcccctt 62
Db 9095 ctactaaatggttcaaacgcttcgcttcgctcactcgggaccggtctaaagccgcccctt 62

QY 63 aaccacacgttaga 76
Db 9155 aaccacacgttaga 9168

RESULT 9
PCT-US99-13295-19
; Sequence 19, Application PC/TUS9913295
; GENERAL INFORMATION:
; APPLICANT: New England Biolabs, Inc.
; APPLICANT: Vaisvila, Romualdas
; APPLICANT: Morgan, Richard D.
; APPLICANT: Raleigh, Elisabeth
; TITLE OF INVENTION: Restriction Enzyme Gene Discovery Method
; FILE REFERENCE: Gene Discovery Method
; CURRENT APPLICATION NUMBER: PCT/US99/13295
; CURRENT FILING DATE: 1999-06-11
; EARLIER APPLICATION NUMBER: 60/089,086
; EARLIER FILING DATE: 1998-06-12
; EARLIER APPLICATION NUMBER: 60/089,101
; EARLIER FILING DATE: 1998-06-12
; NUMBER OF SEQ ID NOS: 130
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
; LENGTH: 77
; TYPE: DNA
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: Genomic DNA of
; OTHER INFORMATION: Pseudomonas Alcaligenes NEB #585 (ATCC 55044)
PCT-US99-13295-19

Query Match          93.2%; Score 71.8; DB 1; Length 77;
Best Local Similarity 97.3%; Pred. No. 4.6e-17;
Matches 73; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ctctacaaatggttcaaacgcttcgcttcgctcactcgggaccggtctaaagccgccc 60
Db 1 ctctacaaatggttcaaacgcttcgcttcgctcactcgggaccggtctaaagccgccc 60

QY 61 ttaacacacgttag 75
Db 61 ttaacacacgttag 75

RESULT 10
US-09-701-626A-19
; Sequence 19, Application US/09701626A
; GENERAL INFORMATION:
; APPLICANT: Vaisvila, Romualdas
; APPLICANT: Morgan, Richard D.
; APPLICANT: Raleigh, Elisabeth
; TITLE OF INVENTION: Restriction Enzyme Gene Discovery Method
; FILE REFERENCE: NEB-165PUS
; CURRENT APPLICATION NUMBER: US/09/701,626A
; CURRENT FILING DATE: 2001-04-16
; PRIOR APPLICATION NUMBER: 60/089,086
; PRIOR FILING DATE: 1998-06-12
; PRIOR APPLICATION NUMBER: 60/089,101
; PRIOR FILING DATE: 1998-06-12
; PRIOR APPLICATION NUMBER: PCT/US99/13295
; PRIOR FILING DATE: 1999-06-11
; NUMBER OF SEQ ID NOS: 130
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
; LENGTH: 77
; TYPE: DNA
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: Genomic DNA of
; OTHER INFORMATION: Pseudomonas Alcaligenes NEB #585 (ATCC 55044)
US-09-701-626A-19

Query Match          93.2%; Score 71.8; DB 28; Length 77;
Best Local Similarity 97.3%; Pred. No. 4.6e-17;
Matches 73; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ctctacaaatggttcaaacgcttcgcttcgctcactcgggaccggtctaaagccgccc 60
Db 1 ctctacaaatggttcaaacgcttcgcttcgctcactcgggaccggtctaaagccgccc 60

QY 61 ttaacacacgttag 75
Db 61 ttaacacacgttag 75

RESULT 11
PCT-US99-13295-63
; Sequence 63, Application PC/TUS9913295
; GENERAL INFORMATION:
; APPLICANT: New England Biolabs, Inc.
; APPLICANT: Vaisvila, Romualdas
; APPLICANT: Morgan, Richard D.
; APPLICANT: Raleigh, Elisabeth
; TITLE OF INVENTION: Restriction Enzyme Gene Discovery Method
; FILE REFERENCE: Gene Discovery Method
; CURRENT APPLICATION NUMBER: PCT/US99/13295
; CURRENT FILING DATE: 1999-06-11
; EARLIER APPLICATION NUMBER: 60/089,086
; EARLIER FILING DATE: 1998-06-12
; EARLIER APPLICATION NUMBER: 60/089,101
; EARLIER FILING DATE: 1998-06-12
; NUMBER OF SEQ ID NOS: 130
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 63
; LENGTH: 77
; TYPE: DNA
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: Genomic DNA of
; OTHER INFORMATION: Pseudomonas Alcaligenes NEB #585 (ATCC 55044)
PCT-US99-13295-63

Query Match          93.2%; Score 71.8; DB 1; Length 77;
Best Local Similarity 97.3%; Pred. No. 4.6e-17;
Matches 73; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ctctacaaatggttcaaacgcttcgcttcgctcactcgggaccggtctaaagccgccc 60
Db 1 ctctacaaatggttcaaacgcttcgcttcgctcactcgggaccggtctaaagccgccc 60

QY 61 ttaacacacgttag 75
Db 61 ttaacacacgttag 75
```

```

Matches 73; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ctctacaaatggttcaaacgcttcgcttcgctcactcgggaccggtctaaagccgccc 60
Db 1 ctctacaaatggttcaaacgcttcgcttcgctcactcgggaccggtctaaagccgccc 60

QY 61 ttaacacacgttag 75
Db 61 ttaacacacgttag 75
```

```

RESULT 10
US-09-701-626A-19
; Sequence 19, Application US/09701626A
; GENERAL INFORMATION:
; APPLICANT: Vaisvila, Romualdas
; APPLICANT: Morgan, Richard D.
; APPLICANT: Raleigh, Elisabeth
; TITLE OF INVENTION: Restriction Enzyme Gene Discovery Method
; FILE REFERENCE: NEB-165PUS
; CURRENT APPLICATION NUMBER: US/09/701,626A
; CURRENT FILING DATE: 2001-04-16
; PRIOR APPLICATION NUMBER: 60/089,086
; PRIOR FILING DATE: 1998-06-12
; PRIOR APPLICATION NUMBER: 60/089,101
; PRIOR FILING DATE: 1998-06-12
; PRIOR APPLICATION NUMBER: PCT/US99/13295
; PRIOR FILING DATE: 1999-06-11
; NUMBER OF SEQ ID NOS: 130
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
; LENGTH: 77
; TYPE: DNA
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: Genomic DNA of
; OTHER INFORMATION: Pseudomonas Alcaligenes NEB #585 (ATCC 55044)
US-09-701-626A-19
```

```

Query Match          93.2%; Score 71.8; DB 28; Length 77;
Best Local Similarity 97.3%; Pred. No. 4.6e-17;
Matches 73; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ctctacaaatggttcaaacgcttcgcttcgctcactcgggaccggtctaaagccgccc 60
Db 1 ctctacaaatggttcaaacgcttcgcttcgctcactcgggaccggtctaaagccgccc 60

QY 61 ttaacacacgttag 75
Db 61 ttaacacacgttag 75

RESULT 11
PCT-US99-13295-63
; Sequence 63, Application PC/TUS9913295
; GENERAL INFORMATION:
; APPLICANT: New England Biolabs, Inc.
; APPLICANT: Vaisvila, Romualdas
; APPLICANT: Morgan, Richard D.
; APPLICANT: Raleigh, Elisabeth
; TITLE OF INVENTION: Restriction Enzyme Gene Discovery Method
; FILE REFERENCE: Gene Discovery Method
; CURRENT APPLICATION NUMBER: PCT/US99/13295
; CURRENT FILING DATE: 1999-06-11
; EARLIER APPLICATION NUMBER: 60/089,086
; EARLIER FILING DATE: 1998-06-12
; EARLIER APPLICATION NUMBER: 60/089,101
; EARLIER FILING DATE: 1998-06-12
; NUMBER OF SEQ ID NOS: 130
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 63
; LENGTH: 77
```

```
; TYPE: DNA
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: Genomic DNA of
; OTHER INFORMATION: Pseudomonas Alcaligenes NEB #585 (ATCC 55044)
; PCT-US99-13295-63

Query Match          92.7%; Score 71.4; DB 1; Length 77;
Best Local Similarity 98.6%; Pred. No. 6.6e-17;
Matches 72; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 ctaacaaatggttcaagcgttcgtctcactcgtcggaacggcgttaaacggccctt 62
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3 ctaacaaatggttcaagcgttcgtctcactcgtcggaacggcgttaaacggccctt 62

QY 63 aaccaaactgttag 75
   ||||| ||||| |||||
Db 63 aaccaaactgttag 75

RESULT 12
PCT-US99-13295-119
; Sequence 119, Application PC/TUS9913295
; GENERAL INFORMATION:
; APPLICANT: New England Biolabs, Inc.
; APPLICANT: Vaisvila, Romualdas
; APPLICANT: Morgan, Richard D.
; APPLICANT: Raleigh, Elisabeth
; TITLE OF INVENTION: Restriction Enzyme Gene Discovery Method
; FILE REFERENCE: Gene Discovery Method
; CURRENT APPLICATION NUMBER: PCT/US99/13295
; EARLIER FILING DATE: 1999-06-11
; EARLIER FILING DATE: 1998-06-12
; EARLIER FILING DATE: 1998-06-12
; EARLIER FILING DATE: 1998-06-12
; NUMBER OF SEQ ID NOS: 130
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 119
; LENGTH: 77
; TYPE: DNA
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: Consensus
; OTHER INFORMATION: Sequence
; PCT-US99-13295-119

Query Match          92.7%; Score 71.4; DB 1; Length 77;
Best Local Similarity 98.6%; Pred. No. 6.6e-17;
Matches 72; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 ctaacaaatggttcaagcgttcgtctcactcgtcggaacggcgttaaacggccctt 62
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3 ctaacaaatggttcaagcgttcgtctcactcgtcggaacggcgttaaacggccctt 62

QY 63 aaccaaactgttag 75
   ||||| ||||| |||||
Db 63 aaccaaactgttag 75

RESULT 13
US-09-701-626A-63
; Sequence 63, Application US/09701626A
; GENERAL INFORMATION:
; APPLICANT: Vaisvila, Romualdas
; APPLICANT: Morgan, Richard D.
; APPLICANT: Raleigh, Elisabeth
; TITLE OF INVENTION: Restriction Enzyme Gene Discovery Method
; FILE REFERENCE: NEB-165PUS
; CURRENT APPLICATION NUMBER: US/09/701,626A
; CURRENT FILING DATE: 2001-04-16
```

```
; PRIOR APPLICATION NUMBER: 60/089,086
; PRIOR FILING DATE: 1998-06-12
; PRIOR APPLICATION NUMBER: 60/089,101
; PRIOR FILING DATE: 1998-06-12
; PRIOR APPLICATION NUMBER: PCT/US99/13295
; PRIOR FILING DATE: 1999-06-11
; NUMBER OF SEQ ID NOS: 130
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 63
; LENGTH: 77
; TYPE: DNA
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: Genomic DNA of
; OTHER INFORMATION: Pseudomonas Alcaligenes NEB #585 (ATCC 55044)
; US-09-701-626A-63

Query Match          92.7%; Score 71.4; DB 28; Length 77;
Best Local Similarity 98.6%; Pred. No. 6.6e-17;
Matches 72; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 ctaacaaatggttcaagcgttcgtctcactcgtcggaacggcgttaaacggccctt 62
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3 ctaacaaatggttcaagcgttcgtctcactcgtcggaacggcgttaaacggccctt 62

QY 63 aaccaaactgttag 75
   ||||| ||||| |||||
Db 63 aaccaaactgttag 75

RESULT 14
US-09-701-626A-119
; Sequence 119, Application US/09701626A
; GENERAL INFORMATION:
; APPLICANT: Vaisvila, Romualdas
; APPLICANT: Morgan, Richard D.
; APPLICANT: Raleigh, Elisabeth
; TITLE OF INVENTION: Restriction Enzyme Gene Discovery Method
; FILE REFERENCE: NEB-165PUS
; CURRENT APPLICATION NUMBER: US/09/701,626A
; CURRENT FILING DATE: 2001-04-16
; PRIOR APPLICATION NUMBER: 60/089,086
; PRIOR FILING DATE: 1998-06-12
; PRIOR APPLICATION NUMBER: 60/089,101
; PRIOR FILING DATE: 1998-06-12
; PRIOR APPLICATION NUMBER: PCT/US99/13295
; PRIOR FILING DATE: 1999-06-11
; NUMBER OF SEQ ID NOS: 130
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 119
; LENGTH: 77
; TYPE: DNA
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: Consensus
; OTHER INFORMATION: Sequence
; US-09-701-626A-119

Query Match          92.7%; Score 71.4; DB 28; Length 77;
Best Local Similarity 98.6%; Pred. No. 6.6e-17;
Matches 72; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 ctaacaaatggttcaagcgttcgtctcactcgtcggaacggcgttaaacggccctt 62
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3 ctaacaaatggttcaagcgttcgtctcactcgtcggaacggcgttaaacggccctt 62

QY 63 aaccaaactgttag 75
   ||||| ||||| |||||
Db 63 aaccaaactgttag 75
```

```

RESULT 15
PCT-US99-13295-65
: Sequence 65, Application PC/TUS9913295
: GENERAL INFORMATION:
: APPLICANT: New England Biolabs, Inc.
: APPLICANT: Valsvila, Romualdas
: APPLICANT: Morgan, Richard D.
: APPLICANT: Raleigh, Elisabeth
: TITLE OF INVENTION: Restriction Enzyme Gene Discovery Method
: FILE REFERENCE: Gene Discovery Method
: CURRENT APPLICATION NUMBER: PCT/US99/13295
: EARLIER FILING DATE: 1999-06-11
: EARLIER FILING DATE: 1998-06-12
: EARLIER FILING DATE: 1998-06-12
: EARLIER FILING DATE: 1998-06-12
: NUMBER OF SEQ ID NOS: 130
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 65
: LENGTH: 77
: TYPE: DNA
: ORGANISM: Unknown
: FEATURE:
: OTHER INFORMATION: Description of Unknown Organism: Genomic DNA of
: OTHER INFORMATION: Pseudomonas Alcaligenes NEB #585 (ATCC 55044)
PCT-US99-13295-65

```

```

Query Match          92.5%; Score 71.2; DB 1; Length 77;
Best Local Similarity 96.1%; Pred. No. 7.8e-17;
Matches 73; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 tctaacaaatggtcaagcgcgttcgtcgtcactcgggaccggtctaaagcgcgcct 61
Db 2 tctaacatggttcaagcgcgttcgtcgtcactcgggaccggtctaaagcgcgcct 61
QY 62 taaccaaacggttagag 77
Db 62 taaccaaacggttagag 77

```

Search completed: August 2, 2001, 17:44:11
Job time: 3062 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 2, 2001, 16:53:10 ; Search time 1350.43 Seconds
(without alignments)
881.954 Million cell updates/sec

Title: US-09-701-626A-78

Perfect score: 77
Sequence: 1 ctctaacaaatggttcaagc.....cccttaacaaacgttagag 77

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues
Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*

- 1: gb_ba1:*
- 2: gb_ba2:*
- 3: gb_ba3:*
- 4: gb_in1:*
- 5: gb_in2:*
- 6: gb_in3:*
- 7: gb_om:*
- 8: gb_ov:*
- 9: gb_pat1:*
- 10: gb_pat2:*
- 11: gb_ph:*
- 12: gb_pl1:*
- 13: gb_pl2:*
- 14: gb_pl3:*
- 15: gb_pl4:*
- 16: em_ba1:*
- 17: em_ba2:*
- 18: em_fun:*
- 19: em_higo_hum:*
- 20: em_higo_inv:*
- 21: em_higo_rod:*
- 22: em_hig_hum1:*
- 23: em_hig_hum2:*
- 24: em_hig_hum3:*
- 25: em_hig_hum4:*
- 26: em_hig_hum5:*
- 27: em_hig_hum6:*
- 28: em_hig_hum7:*
- 29: em_hig_hum8:*
- 30: em_hig_inv1:*
- 31: em_hig_inv2:*
- 32: em_hig_other:*
- 33: em_hig_rod:*
- 34: em_hum1:*
- 35: em_hum2:*
- 36: em_hum3:*
- 37: em_hum4:*
- 38: em_hum5:*
- 39: em_hum6:*
- 40: em_hum7:*
- 41: em_in:*
- 42: em_om:*
- 43: em_or:*

44: em_ov:*

45: em_pat:*

46: em_ph:*

47: em_pl:*

48: em_ro:*

49: em_sts:*

50: em_sy:*

51: em_un:*

52: em_v1:*

53: gb_stsl:*

54: gb_sts2:*

55: gb_sts3:*

56: gb_sy:*

57: gb_un:*

58: gb_v11:*

59: gb_v12:*

60: gb_htg1:*

61: gb_htg2:*

62: gb_htg3:*

63: gb_htg4:*

64: gb_htg5:*

65: gb_htg6:*

66: gb_htg7:*

67: gb_htg8:*

68: gb_htg9:*

69: gb_htg10:*

70: gb_htg11:*

71: gb_htg12:*

72: gb_htg13:*

73: gb_htg14:*

74: gb_htg15:*

75: gb_htg16:*

76: gb_htg17:*

77: gb_htg18:*

78: gb_htg19:*

79: gb_htg20:*

80: gb_htg21:*

81: gb_htg22:*

82: gb_htg23:*

83: gb_htg24:*

84: gb_htg25:*

85: gb_pr1:*

86: gb_pr2:*

87: gb_pr3:*

88: gb_pr4:*

89: gb_pr5:*

90: gb_pr6:*

91: gb_pr7:*

92: gb_pr8:*

93: gb_pr9:*

94: gb_rol:*

95: gb_rod:*

96: gb_in4:*

97: gb_pr10:*

98: em_ba3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	40.4	52.5	9581	3	PAE223604	AJ223604 Pseudomon
2	30.2	39.2	174732	62	AC012160	AC012160 Drosophil
3	26.8	34.8	179936	69	AC025491	AC025491 Homo sapi
4	26.6	34.5	11014	3	MTPFZ1A	X68367 M.thermofo
5	26.4	34.3	12249	1	AE000501	AE000501 Escherich
6	26.4	34.3	28446	66	AC020814	AC020814 Mus muscu
7	26.4	34.3	160312	62	AC011073	AC011073 Homo sapi
8	26.4	34.3	181215	86	AC006461	AC006461 Homo sapi

```

9      26.4 34.3 338534 2  ECOW93
c 10    26 33.8 181931 79  AL353773
c 11    26 33.8 195032 86  AC006239
c 12    25.8 33.5 195000 43  AP000396
c 13    25.2 32.7 2051 2  AY007784
c 14    25.2 32.7 10521 65  AC020168
c 15    25.2 32.7 37027 89  AL135844
c 16    25.2 32.7 142229 30  AC007753
c 17    25.2 32.7 184272 4  AC009211
c 18    25.2 32.7 196928 82  AP000847
c 19    25.2 32.7 225668 5  AE003699
c 20    25 32.5 199380 72  AC048379
c 21    25 32.5 210000 60  AC004555
c 22    24.8 32.2 828 2  AF217619
c 23    24.8 32.2 6502 85  AB038980S2
c 24    24.8 32.2 10945 1  AE006262
c 25    24.8 32.2 14466 1  AE005662
c 26    24.8 32.2 151840 69  AC026226
c 27    24.8 32.2 168764 2  AP002569
c 28    24.8 32.2 176919 67  AC023169
c 29    24.8 32.2 182178 60  AC007236
c 30    24.8 32.2 182309 75  AC074016
c 31    24.4 31.7 87258 88  AF091512
c 32    24.4 31.7 110000 84  LMFLCHR26_3
c 33    24.4 31.7 141832 91  AP000432
c 34    24.4 31.7 151097 76  AC083846
c 35    24.4 31.7 152044 73  AC088255
c 36    24.4 31.7 168700 61  AC009115
c 37    24.4 31.7 196055 69  AC025513
c 38    24.4 31.7 239566 93  HSAJ03147
c 39    24.4 31.7 340000 91  AP001670
c 40    24.2 31.4 473 1  ABA251519
c 41    24.2 31.4 3407 3  RSP243600
c 42    24.2 31.4 118795 60  AC008618
c 43    24.2 31.4 167086 86  AC008565
c 44    24.2 31.4 171909 70  AC026470
c 45    24.2 31.4 173396 80  AL390236

```

ALIGNMENTS

```

RESULT 1
PAE223604      PAE223604      9581 bp      BCT      19-JUL-2000
LOCUS          Pseudomonas aeruginosa integron In 101 DNA.
DEFINITION     Pseudomonas aeruginosa
ACCESSION      AJ223604
VERSION         AJ223604.1 GI:4210822
KEYWORDS       Pseudomonas aeruginosa.
SOURCE         Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
                Pseudomonas.
REFERENCE      1 (bases 1 to 9581)
AUTHORS        Laraki,N., Galleni,M., Thamm,I., Riccio,M.L., Amicosante,G.,
                Frere,J.M. and Rossolini,G.M.
TITLE          Structure of In31, a blaIMP-containing Pseudomonas aeruginosa
                integron phyletically related to In5, which carries an unusual
                array of gene cassettes
JOURNAL        Antimicrob. Agents Chemother. 43 (4), 890-901 (1999)
MEDLINE        99216901
REFERENCE      2 (bases 1 to 9581)
AUTHORS        Galleni,M.G.
TITLE          Direct Submission
JOURNAL        Submitted (20-JAN-1998) Galleni M.G., Centre for Protein
                Engineering (CIP), University of Liege, B6 Sart Tilman, Liege,
                Liege B4000, Belgium
FEATURES       Location/Qualifiers
                1..9581
                /organism="Pseudomonas aeruginosa"
                /db_xref="taxon:287"
                117..140
                /note="IRI"
repeat_region 117..140
source

```

```

/rpt_type-INVERTED
117..1467
/note="5' conserved element"
complement(319..1332)
/gene="intI1"
complement(319..1332)
/gene="intI1"
/codon_start=1
/transl_table=11
/protein_id="CA11470.1"
/db_xref="GI:4210823"
/db_xref="SPTREMBL:O9R819"
/translacion="MKTATAPLPPLRSVKVLDQLRRIRVLYSLRTEQAVVHVRAP
IRHGVRHPATLGSSEVEAFSLNANERKYSVSTHRQALAAALLFYKVLCTDLPWLQ
EIGRPRSRRLPVVLPDVEVRIILGFLLEGHRLFAQLLYGTGMRISGLQRLKDLF
DHGTIIVREGSKDRALMPLPESLAPTLRQLSRARAWLKDQAEGRSGVALPALER
KYPRAGSHFWFWFAQHTSTDPSCGVRRHHMYDQTFQRAFRAYEQAGITKPATP
HTLRHSFATALRSYDITRVQDLLGHSDVSTTMYTHVLKVGAGVRSPLDALPPLT
SER"
1468..2346
/gene="bla-imp"
1468..2346
/gene="bla-imp"
1485..2225
/gene="bla-imp"
/codon_start=1
/transl_table=11
/protein_id="CA11471.1"
/db_xref="GI:4210824"
/translacion="MSKLSVFFIFLFCSIATAESLPDLKTEKIDEGVYVHTSPEEVN
GWGVPHGIVLVNAPAYLIDTPPTAKDTKLVTFWVERGYKIKGSISSHFHSDSTG
GIWLNRSIPTVASELTNELLKDKGKQVATNSFGVNYHLVKNKIEVYPCGHTPD
NVVWLPERKILGCGCFIKPIYGLNLDGANIEAWPKSAKLLKSKYGRKAKLVVPSHSEV
GDASLLKLTLEQAVKGLNESKPKSPSN"
2262..2346
/gene="bla-imp"
/note="59 bp element"
2347..2984
/gene="aacA4"
2347..2984
/gene="aacA4"
2371..2925
/gene="aacA4"
/codon_start=1
/transl_table=11
/protein_id="CA11472.1"
/db_xref="GI:4210825"
/db_xref="SPTREMBL:Q51410"
/translacion="MTNSTDSVTLRMTLMEHDLAMLYEWLNKSHIVEWMGGEARPTLA
DVQBYLPYSLAQESVTPYIAMLNGEPYGAQSVYALGSGDGMWEEETDPCVRGIDQL
LANASOLGKLGTKLVRALVELLFNDPEVTKIQTPSPSNLRAIRCYEKAGFERQGTV
TTPDGPAYVVOIRQAFERTSDA"
2920..2984
/gene="aacA4"
/note="59 bp element"
2985..3715
/gene="catB6"
2985..3715
/gene="catB6"
3018..3650
/gene="catB6"
/codon_start=1
/transl_table=11
/protein_id="CA11473.1"
/db_xref="GI:4210826"
/db_xref="SPTREMBL:O9R818"
/translacion="MENYFDSPPFKLLSBOVTNRNKKVGRYSYSGYHGHSDDDCA
RYLLPDRDDVDKLIIGSFCISGASFIAGNQGHRHDWTSFPPFQMEAPFSSST
DAFOKAGDEITVNDVWIGSEAMIPGKIGDGAIGVRSVLTVDPEVPTIIGNPAKQ
IKRFSDEEISLLMEMEWNNWPLDKIKTAPLLCCSDIFGLHRHWRGIATV"
3645..3715
/gene="catB6"
/note="59 bp element"

```

misc_feature

```

mRNA      3716..4404
/gene="orf1"
gene       3716..4404
/gene="orf1"
CDS       3727..4341
/gene="orf1"
/codon_start=1
/transl_table=11
/protein_id="CAAL1474.1"
/db_xref="GI:4210827"
/db_xref="SPTREMBL:Q9ZEHL"
/translation="MDVDLSIGIOIVQEALEKGLIFNEKTKTDDFNACDHLVLL
VDAFCDFRSGMTSVFLSITAIEVAKAEAGHEFNLRESSLYFSNENGQFTVPANV
MAILTPVMSKRLEALCKEACALLKEAAGHEFNLRESSLYFSNENGQFTVPANV
SONRAKELLALAEADRRVLGYTTHNTGILEAKINEIFSVAHS"
4335..4404
/gene="orf1"
/misc_feature
/note="59 bp element"
mRNA      4405..4940
/gene="gacE2"
gene       4405..4940
/gene="gacE2"
gene       4507..4839
/gene="gacE2"
CDS       4507..4839
/gene="gacE2"
/codon_start=1
/transl_table=11
/protein_id="CAAL1475.1"
/db_xref="GI:4210828"
/db_xref="SPTREMBL:Q9ZEHO"
/translation="MKNLFLATLAIFFEVIATLSALKSSSEGTFLVPSFIVAGYAAAF
YFLSLTKSIPVGIAAYAVWSGLGIVLTAIAWVLHGQKLDMMGFVGVGFIISGVAVLN
LTKASVH"
4849..4936
/misc_feature
/gene="gacE2"
/note="59 bp element"
misc_feature 4941..6961
/note="3' conserved element"
gene       5045..5392
/gene="gacDelta 1"
CDS       5045..5392
/gene="gacDelta 1"
/codon_start=1
/transl_table=11
/protein_id="CAAL1476.1"
/db_xref="GI:4210829"
/translation="MKGNLFLVIAIVGEVIATLSALKSSSEGTFLKAPSAAVVIIGYIAP
YFLSLVLSIPVGAYAVWSGLGVITIAIAWLLHGQKLDMMGFVGVMGLIIIAFLIAR
SPSWKSLRRPTPW"
5386..6225
/gene="sul1"
CDS       5386..6225
/gene="sul1"
/codon_start=1
/transl_table=11
/protein_id="CAAL1477.1"
/db_xref="GI:4210830"
/translation="MVTVEGLNLTEDSFFDESRRLDPAGAVTAIEMLRVGSDDVDV
GPAASHPARVPSPAEIRRIAPLDLALSDOMHRYISIDSFOPETORYALKRGVGLND
IGQFDPALYDIAEADGRLVVMHSAQRDGIATRGHLRPELDALDEIVRFFARVSAL
RRSGVADRLDPMGFFFLPAPETSLHVLNLOKLSALGSLGLVSVSRKSLFGAT
VGLPVKIDGPSLAAELHAIGNGADYVRTHAPGDLRSATITFSETLAKFRSDARDGL
DHA"
6353..6853
/gene="orf5"
CDS       6353..6853
/gene="orf5"
/codon_start=1
/transl_table=11
/protein_id="CAAL1478.1"
/db_xref="GI:4210831"
/db_xref="SPTREMBL:P75025"

```

```

/translation="MDEEPNPNVACSGDIDFVRLHMDHDAAMWSAKGTPANDVARI
DRTPAFTFVLRSELLVASCSDGIVGCGCTLSAEDPEFPWFDALKEAALYHLKLAARRTHA
GRGVSSALIEACRHAARTQCAKLRDCHPNRLGLYERLGFTHVDYTFNPGWDPFTFAIE
RLELEI"
6962..7320
/note="359 bp identical to In5"
/misc_feature
gene       complement(7311..7736)
/gene="tniB delta2"
CDS       complement(7311..7736)
/gene="tniB delta2"
/codon_start=1
/transl_table=11
/protein_id="CAAL1479.1"
/db_xref="GI:4210832"
/db_xref="SPTREMBL:Q9R817"
/translation="MDEYPIIDLSHLLPAAQGLARLPADERIORLADRMIYIPRAVE
ALNRLEALYAMPNKRPNLLVCPNTNGKSMIVKEFRTHPASSDADQEHIPVLVWQ
MPSEPSVIRFYVALLAAMGAPLRPRRLPEMEQLALAW"
complement(7739..9418)
/misc_feature
gene       complement(7739..9418)
/gene="tniA"
CDS       complement(7739..9418)
/gene="tniA"
/codon_start=1
/transl_table=11
/protein_id="CAAL1480.1"
Query Match      52.5%; Score 40.4; DB 3; Length 9581;
Best Local Similarity 89.2%; Pred. No. 0.00029;
Matches 66; Conservative 0; Mismatches 6; Indels 2; Gaps 2;
QY 3 ctaacaatggttcacgcgttcctgcctcactcggcgacgcgcgcgcgcctt 62
      CTTTCTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT
Db 4338 TTAACAATGTTCAATCGCTCGCTCGCTCGCT-GGACGGGCTAAAGCC-GCCCTT 4395
      TTTTCTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT
QY 63 aaccaaagcgttaga 76
      TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT
Db 4396 ACCCAACGTTAGA 4409
      TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT
RESULT 2
AC012160/c
LOCUS
DEFINITION
AC012160 174732 bp DNA HTG 24-NOV-1999
Drosophila melanogaster chromosome X clone BACR06G02 (D1110)
RPCI-98 06.G.2 map 15E-15E strain Y; cn bw sp, *** SEQUENCING IN
PROGRESS ***, 93 unordered pieces.
AC012160
VERSION
AC012160.5 GI:6466928
HTG: HTGS_PHASE1.
SOURCE
fruit fly.
ORGANISM
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 174732)
Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
Hinkle,A., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K.,
Kearney,L., Lee,B., Lewis,S., Li,P., Ling,H., Moshrefi,A.R.,
Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B.,
Richards,S., Sethi,H., Svirskas,R.R., Wan,K.H., Webster,D.,
Woolley,P., Yang,S., Yee,M., Yu,C. and Rubin,G.M.
Sequencing of Drosophila melanogaster
Unpublished
2 (bases 1 to 174732)
Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazdd,P.,
Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and

```

Rubin, G.M.
Direct Submission
Submitted (21-OCT-1999), Drosophila Genome Center, Lawrence Berkeley
Laboratory, MS 64-121, Berkeley, CA 94720, USA
On Nov 24, 1999 this sequence version replaced gl:6226995.
For further information about this sequence, please visit our location
and relationship to other sequences, including its location
archive web site (<http://www.fruitfly.org/sequence/>) or send email
to bdg@fruitfly.berkeley.edu. All contigs in this submission meet
the following cutoffs: length >= 200 bases.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 93 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 572: contig of 572 bp in length
573 652: gap of unknown length
574 1157: contig of 505 bp in length
575 1158 1237: gap of unknown length
576 1238 1708: contig of 471 bp in length
577 1709 1788: gap of unknown length
578 1789 2248: contig of 461 bp in length
579 2249 2329: gap of unknown length
580 2330 3509: contig of 1180 bp in length
581 3510 3589: gap of unknown length
582 3590 4696: contig of 1107 bp in length
583 4697 4776: gap of unknown length
584 4777 5387: contig of 611 bp in length
585 5388 5467: gap of unknown length
586 5468 5971: contig of 504 bp in length
587 5972 6051: gap of unknown length
588 6052 6679: contig of 628 bp in length
589 6680 6759: gap of unknown length
590 6760 7223: contig of 464 bp in length
591 7224 7303: gap of unknown length
592 7304 7890: contig of 587 bp in length
593 7891 7970: gap of unknown length
594 7971 8553: contig of 583 bp in length
595 8554 8633: gap of unknown length
596 8634 9329: contig of 696 bp in length
597 9330 9409: gap of unknown length
598 9410 10185: contig of 776 bp in length
599 10186 10265: gap of unknown length
600 10266 11634: contig of 1369 bp in length
601 11635 11714: gap of unknown length
602 11715 12478: contig of 764 bp in length
603 12479 12558: gap of unknown length
604 12559 14183: contig of 1625 bp in length
605 14184 14263: gap of unknown length
606 14264 15048: contig of 785 bp in length
607 15049 15128: gap of unknown length
608 15129 16098: contig of 968 bp in length
609 16099 16177: gap of unknown length
610 16178 17599: contig of 1423 bp in length
611 17600 17680 17679: gap of unknown length
612 17681 19313: contig of 1634 bp in length
613 19314 19393: gap of unknown length
614 19394 20834: contig of 1441 bp in length
615 20835 20914: gap of unknown length
616 20915 22432: contig of 1518 bp in length
617 22433 22512: gap of unknown length
618 22513 24093: contig of 1581 bp in length
619 24094 24173: gap of unknown length
620 24174 25950: contig of 1777 bp in length
621 25951 26030: gap of unknown length
622 26031 27724 27803: contig of 1693 bp in length
623 27725 29349: gap of unknown length
624 29350 30829: gap of unknown length
625 30830 30909: gap of unknown length
626 30910 32162: contig of 1253 bp in length
627 32163 32242: gap of unknown length
628 32243 33213: contig of 971 bp in length
629 33214 33293: gap of unknown length
630 33294 34919: contig of 1626 bp in length
631 34920 34999: gap of unknown length
632 35000 37862: contig of 2783 bp in length
633 37863 39272: contig of 1410 bp in length
634 39273 39352: gap of unknown length
635 39353 40852: contig of 1500 bp in length
636 40853 40932: gap of unknown length
637 40933 43628: contig of 2696 bp in length
638 43629 43708: gap of unknown length
639 43709 45751: contig of 2043 bp in length
640 45752 45831: gap of unknown length
641 45832 48055: contig of 2224 bp in length
642 48056 48135: gap of unknown length
643 48136 49916: contig of 1781 bp in length
644 49917 49996: gap of unknown length
645 49997 53377: contig of 3381 bp in length
646 53378 53457: gap of unknown length
647 53458 57017: contig of 3360 bp in length
648 57018 57097: gap of unknown length
649 57098 59418: contig of 2320 bp in length
650 59419 59498: gap of unknown length
651 59499 63559: contig of 4062 bp in length
652 63560 63639: gap of unknown length
653 63640 68049: contig of 4410 bp in length
654 68050 68129: gap of unknown length
655 68130 73531: contig of 5402 bp in length
656 73532 73611: gap of unknown length
657 73612 79226: contig of 5615 bp in length
658 79227 79306: gap of unknown length
659 79307 84116: contig of 4810 bp in length
660 84117 91839: contig of 7643 bp in length
661 91840 91919: gap of unknown length
662 91920 98216: contig of 6297 bp in length
663 98217 98296: gap of unknown length
664 98297 106059: contig of 7763 bp in length
665 106060 106139: gap of unknown length
666 106140 112697: contig of 6558 bp in length
667 112698 112777: gap of unknown length
668 112778 120287: contig of 7510 bp in length
669 120288 120367: gap of unknown length
670 120368 128442: contig of 8075 bp in length
671 128443 128522: gap of unknown length
672 128523 139563 139642: gap of unknown length
673 139564 150197: contig of 11040 bp in length
674 150198 150277: gap of unknown length
675 150278 150656: contig of 379 bp in length
676 150657 150736: gap of unknown length
677 150737 151120: contig of 384 bp in length
678 151121 151200: gap of unknown length
679 151201 151836: contig of 636 bp in length
680 151837 151916: gap of unknown length
681 151917 152572: contig of 656 bp in length
682 152573 152652: gap of unknown length
683 152653 153265: contig of 613 bp in length
684 153266 153345: gap of unknown length
685 153346 153842: contig of 497 bp in length
686 153843 153922: gap of unknown length
687 153923 155167: contig of 1245 bp in length
688 155168 155247: gap of unknown length
689 155248 155865: contig of 618 bp in length
690 155866 155945: gap of unknown length
691 155946 156581: contig of 636 bp in length
692 156582 156662: gap of unknown length
693 156663 157211: gap of unknown length

```

* 157212 157702: contig of 491 bp in length
* 157703 157782: gap of unknown length
* 158312 158312: contig of 530 bp in length
* 158313 158392: gap of unknown length
* 158393 159013: contig of 621 bp in length
* 159014 159093: gap of unknown length
* 159094 159952: contig of 859 bp in length
* 159953 160032: gap of unknown length
* 160033 160835: contig of 803 bp in length
* 160836 160915: gap of unknown length
* 160916 161617: contig of 702 bp in length
* 161618 161697: gap of unknown length
* 161698 162146: contig of 449 bp in length
* 162147 162226: gap of unknown length
* 162227 162842: contig of 616 bp in length
* 162843 162942: gap of unknown length
* 162943 163348: contig of 426 bp in length
* 163349 163428: gap of unknown length
* 163429 163916: contig of 488 bp in length
* 163917 163996: gap of unknown length
* 163997 164860: contig of 864 bp in length
* 164861 164940: gap of unknown length

Query Match          39.2%  Score 30.2:  DB 62:  Length 174732;
Best Local Similarity 69.5%  Pred. No. 1 4:
Matches 41:  Conservative 0:  Mismatches 18:  Indels 0:  Gaps 0:

Oy 13 gttcaagcgttccttcgtcactgcgggacgcggtctaaagccggcccttaaccacaaacg 71
||||| ||||| || |||| ||||| || |||| || |||| ||||| ||||| ||||| |||||
Db 18837 GTTCAGCGGTGCTCGCTCTCCGCGGGGACCAACACCGCCACCTTAACCAAGG 18779

RESULT 3
AC025491
LOCUS          AC025491 179936 bp DNA HTG 28-MAR-2000
DEFINITION    Homo sapiens chromosome 6 clone RP11-196B15 map 6, WORKING DRAFT
SEQUENCE     29 unordered pieces.
ACCESSION     AC025491
VERSION       AC025491.2 GI:7331526
KEYWORDS      HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE        human.
ORGANISM      Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE     1 (bases 1 to 179936)
AUTHORS       Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
               Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
               Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G.,
               Campoliano,A., Castie,A., Choepel,Y., Colangelo,M., Collins,S.,
               Collymore,A., Cooke,P., DeAureliano,K., Dewar,K., Diaz,J.S.,
               Dodge,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,
               Galagan,J., Gardyna,S., Glnde,S., Goyette,M., Graham,L.,
               Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
               Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
               Klein,J., Lakoque,K., Lamazares,R., Landers,T., Lehocsky,J.,
               Levine,R., Liu,G., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
               McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
               Melidrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
               Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
               O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
               Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
               Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
               Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
               Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
               Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
               Young,G., Zainoun,J., Zimmer,A. and Zody,M.
               Direct Submission
TITLE          Submitted (09-MAR-2000) Whitehead Institute/MIT Center for Genome
JOURNAL        Research, 320 Charles Street, Cambridge, MA 02141, USA

```

COMMENT

On Mar 28, 2000 this sequence version replaced gi:7212075.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L7649
Center clone name: 196_B_15
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 163781 bases at least Q40
Consensus quality: 171238 bases at least Q30
Consensus quality: 174757 bases at least Q20
Insert size: 185000; agarose-fp
Insert size: 177136; sum-of-contigs
Quality coverage: 3.8 in Q20 bases; agarose-fp
Quality coverage: 3.9 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 29 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1579: contig of 1579 bp in length
* 1580 1679: gap of 100 bp
* 1680 3543: contig of 1864 bp in length
* 3544 3643: gap of 100 bp
* 3644 5682: contig of 2039 bp in length
* 5683 5782: gap of 100 bp
* 5783 7667: contig of 1885 bp in length
* 7668 7767: gap of 100 bp
* 7768 10500: contig of 2733 bp in length
* 10501 10600: gap of 100 bp
* 10601 12890: contig of 2290 bp in length
* 12891 12990: gap of 100 bp
* 12991 14880: contig of 1890 bp in length
* 14881 14980: gap of 100 bp
* 14981 17380: contig of 2400 bp in length
* 17381 17480: gap of 100 bp
* 17481 20807: contig of 3327 bp in length
* 20808 20907: gap of 100 bp
* 20908 23928: contig of 3021 bp in length
* 23929 24048: gap of 100 bp
* 24049 27543: contig of 3515 bp in length
* 27544 27643: gap of 100 bp
* 27644 31065: contig of 3422 bp in length
* 31066 31165: gap of 100 bp
* 31166 35411: contig of 4246 bp in length
* 35412 35511: gap of 100 bp
* 35512 39412: contig of 3901 bp in length
* 39413 39512: gap of 100 bp
* 39513 43164: contig of 3652 bp in length
* 43165 43264: gap of 100 bp
* 43265 46755: contig of 3491 bp in length
* 46756 46855: gap of 100 bp
* 46856 53117: contig of 6262 bp in length
* 53118 53217: gap of 100 bp
* 53218 57708: contig of 4491 bp in length
* 57709 57808: gap of 100 bp
* 57809 62533: contig of 4725 bp in length
* 62534 62633: gap of 100 bp
* 62634 68232: contig of 5599 bp in length
* 68233 68332: gap of 100 bp
* 68333 75673: contig of 7341 bp in length


```

/note="ORF2'"
/codon_start=1
/transl_table=1
/protein_id="CAA48441.1"
/db_xref="GI:44649"
/db_xref="SWISS-PROT:P29572"
/translation="MNSKSLFPGHTDESSQNKGRYRAIGVMVSHPAEFGSELEGDINSI
LDSVDSGLKRRLEKNKKIDIFRCRAIEKIVDFLELSNLGNPPVRVDILRWDIIDS
RHSIOGRDQNONQRMYYHLFSNVISKRWPSGDWCFPDETGSVDWEVAFDLDIGGS
KVDLKEQFINLISINEVDSKDNVLQVADFFAGLSVKEKFGLYVDWNTYECGQORLV
PVNVEFSKKDRKRFEILSYFNKRCKELKMGVSLDTSGLWTPKPANSINFHWEPQS
DADKAPIR"
3455..4108
/note="ORF3'"
/codon_start=1
/transl_table=1
/protein_id="CAA48442.1"
/db_xref="GI:44650"
/db_xref="SWISS-PROT:P29574"
/translation="MKLNKIMILIMHRYFLRVMNLRQKIVALEEGKSVAIKYODV
RDYLDLTKHGFVLEPHINPAKETAAELADLGNLAKSTIYSKYTNEIVRILKRSK
NRNVLVFNQFLQSLKNTARVLLDMEDVOVLCSIRGRPOGGRLLKRWITLSDRSK
EVTDKIPLVFAESFIALITFVKAGSAIYNRNHDFVFLSAAIFVGISVGRTLWIS"
4225..4929
/note="ORF4'"
/codon_start=1
/transl_table=1
/protein_id="CAA48443.1"
/db_xref="GI:44651"
/db_xref="SWISS-PROT:P29576"
/translation="MYTARPIFGCGGILMLSPERRSFTILFYVDNSFICSGRLCRKG
GVNEKHQILNHFVLIIPFVLLYFDYDTGTAAMLSGVSSHILDFDMTGCPPFYIY
KGRYRVDRHKNGSGPREKRALTYIGILAAILLIVAPSPFATSISOMKSGSGSAS
NRRTDINVNFRCNGDTWIHPYNGSIFIDYVSDGDSRVYRYGIRSGGQGHKLLKD
SNTTENTTKQNETGG"
5295..6794
/note="ORF5'"
/codon_start=1
/transl_table=1
/protein_id="CAA48444.1"
/db_xref="GI:44652"
/db_xref="SWISS-PROT:P29579"
/translation="MRNGTLLIALLAVALFVLAGSSAADVGVELDKNTKPVYNSTLR
VKVITAKAGNOVNQNAVIVKVPBGLVLQDYTAGGYDLETGTWEIGDIPAYEERSLT
LVCLNTGNVTYTAANNIDGNDENPANNALQFRVGYDLNLTNVTSSQSARLGD
VTNFKLKNRPHAANNINANFESGCLAQSYTYGYTGYDDVARWILETLDTGREA
TLTVCLVNRGTGLSDSY/RE/VEVDEGDVYNNIARASVAVRGTDLDLDLSVSKPRAY
QGDVVNVRRVKNNGPAAQV/RYNQLPANLQVQVQVDRGYTSYNGVWVIGDLADNE
AALLNITRVYVSNAGFNTATVSPADISDNPVNDDTALVSAAIEPKALKVRIKNS
AVTIRLLVYINIDHGIRTKRTNYFLKNGLSRDLSLGYFQLGTSALFKQYTYNTYNR
PRTVSYENTYNTASVITQVRVNVSGVGRQKAPVVRITALLDENGTSLQ"
6894..7262
/note="ORF6'"
/codon_start=1
/transl_table=1
/protein_id="CAA48445.1"
/db_xref="GI:44653"
/db_xref="SWISS-PROT:P29581"
/translation="MRQIAMRKEVYSGHGDVRMVEVPGMKVIAIGDGVNYLFMLLP
FRDFKUKRDKDWMFLFKLSTYSFOARPPMWTFDKMLSFYPLHLEEAEGHFELDIRNSR
GLMFSFDITYSEQLQORLVV"
7539..8147
/note="ORFX"
/codon_start=1
/transl_table=1
/protein_id="CAA48446.1"
/db_xref="GI:809724"
/db_xref="SWISS-PROT:P29566"
/translation="MIDNFKIEILDFKDELKKITGKEVEFPKYTQIINLANQNAQGT
RPRVVGMSDLIEHCPDQSVGECHKWYLEKSDRIEKATKKISKMTENKKAAMELIDE
EMIKKWEDVLYIKTAEGLLIQILIKLTAEEAGLEWRLATSKESKNTDGFISTGPY
SIKPMSTYENYVREIDQITFYFKKPKNSRYLYIYHNLI"
complement(8136..9203)

```

[illegible]

Query Match 34.5%; Score 26.6; DB 3; Length 11014;

Qy 14 ttcaagcgggttcgcttcgctcactcgggacggcgtaaacgcccctt 62
||| ||||| ||||| ||| ||| ||| |||
Db 3111 TTCGGGGCGTTCGCTTCGCTCACTCTCGAATTCTCTAAACAGCGCTT 3159

RESULT 5
AE000501

LOCUS	AE000501	12249 bp	DNA	BCT	01-DEC-2000
DEFINITION	Escherichia coli K12 MG1655 section 391 of 400 of the complete genome.				
ACCESSION	AE000501 U00096				
VERSION	AE000501.1 GI:2367372				
KEYWORDS					
SOURCE	Escherichia coli K12.				
ORGANISM	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.				
REFERENCE	1 (bases 1 to 12249)				
AUTHORS	Blattner,F.R., Plunkett,G. III, Bloch,C.A., Perna,N.T., Burland,V., Riley,M., Collado-Vides,J., Glasner,J.D., Rode,C.K., Mayhew,G.P., Gregor,J., Davis,N.W., Kirkpatrick,H.A., Goeden,M.A., Rose,D.J., Mau,B. and Shao,Y.				
TITLE	The complete genome sequence of Escherichia coli K-12				
JOURNAL	Science 277 (5331), 1453-1474 (1997)				
MEDLINE	97426617				
PUBMED	9278503				
REFERENCE	2 (bases 1 to 12249)				
AUTHORS	Blattner,F.R.				
TITLE	Direct Submission				
JOURNAL	Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.				
	Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459				
REFERENCE	3 (bases 1 to 12249)				
AUTHORS	Blattner,F.R.				
TITLE	Direct Submission				
JOURNAL	Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.				
	Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459				
REFERENCE	4 (bases 1 to 12249)				
AUTHORS	Plunkett,G. III.				
TITLE	Direct Submission				
JOURNAL	Submitted (13-OCT-1998) Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA				
COMMENT	On Sep 9, 1997 this sequence version replaced gi:1790756. This sequence was determined by the E. coli Genome Project at the University of Wisconsin-Madison (Frederick R. Blattner, director). Supported by NIH grants HG00301 and HG01428 (from the Human Genome Project and NCHGR). The entire sequence was independently determined from E. coli K12 strain MG1655. Predicted open reading frames were determined using GeneMark software, kindly supplied by Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA, 30332 (e-mail: mark@amber.gatech.edu). Open reading frames that have been correlated with genetic loci are being annotated with CG Site Nos., unique ID nos. for the genes in the E. coli Genetic Stock Center (CGSC) database at Yale University, kindly supplied by Mary Berlyn. A public version of the database is accessible (http://cgsc.biology.yale.edu). Annotation of the genome is an ongoing task whose goal is to make the genome sequence more useful by correlating it with other data. Comments to the authors are appreciated. Updated information will be available at the E. coli Genome Project's World Wide Web site (http://www.genetics.wisc.edu). *** The E. coli K12 sequence and its annotations are periodically updated; this is version M54. No sequence changes. Annotation updates: updated gene identifications and products; all new functional assignments courtesy of Monica Riley; added promoters, protein binding sites, and repeated sequences described in reference 1. The unique numeric identifiers beginning with a lowercase 'b' assigned to each gene (protein- or RNA-encoding) are now designated as gene synonyms instead of labels. This should allow them to be searched for in Entrez as gene names.				
FEATURES	source	Location/Qualifiers			
		1. .12249			
		/organism="Escherichia coli K12"			
		/strain="K12"			
		/sub_strain="MG1655"			
		/db_xref="taxon:83333"			
		complement(96..902)			
	gene				
		/gene="sgcQ"			
		complement(96..902)			
		/gene="sgcQ"			
		/function="putative enzyme; Not classified"			
		/note="f268a; formerly designated yjhm"			
		/codon_start=1			
		/transl_table=11			
		/product="putative nucleoside triphosphatase"			
		/protein_id="AAC77259.1"			
		/db_xref="GI:1790757"			
		/translation="MSWLKEVIGTEKAVIAMCHLRALPGDPSFDQLGMNVIDKAMD DMLAQNGVDVAFMSNEFSLPYLTVKRPETTAAMARIIGLMSDIRIPFGVNVLMDDP VASFDLMAATKFIREFTGAYASDFGVMDTNYGETIRHQHRRIGAGGVKTLFNITVPE AAVYLNDRDICSIAKSTVNNHPDALCVSLGTACTRTDSALLKRVKTVPPDVTVLANT GCVLNVPEQLSIADGCVTATTFKKDGVFANFVQDARYSQFMKVVHHR"			
		complement(905..931)			
	promoter	/note="factor Sigma70; predicted +1 start at 4526481"			
		complement(915..2228)			
	gene	/gene="sgcC"			
		/note="b4304"			
		complement(915..2228)			
	CDS	/gene="sgcC"			
		/function="putative transport; Not classified"			
		/note="f437; formerly designated yjhn"			
		/codon_start=1			
		/transl_table=11			
		/product="putative Pts system enzyme IIC component"			
		/protein_id="AAC77260.1"			
		/db_xref="GI:1790758"			
		/translation="MFYILSLGGTVFVPIIMVIGLIFRIPWLOAIKAGVTVGIGFV GMGLIVMAIDSLSPPIKMIERFGLAHFVDVGAGAGVGYATAGAMILIPVIFLL NVAMLVTRLTDMNVDIYWHYAITGTVOLMTGSLYGVLGATCHAALSLLKMDWT AKRVONIVGLRISITPOGYSVSPFLVLLDAIVEKIPFMKGRNIDAEIOKRYGMVG DPVILVGLIFGLIAGAGEGKGCASLMTVAAMVLPFRMIRLIVELLLPSIDGARK FFQYFKGREVYIGLDTAVTLGHPHTIAGULLIPIMLILASILDPGNVLPLDLPVA PFLTCMATVIRHGDVLRVTLISGVIVMTVLLIATQFAPYFTEMLKGGFSFAGESAQI SALSVMNFMGSIISLMSLGIIGVVAVGVIVASVVVFLRKRKRELSE"			
		complement(923..950)			
	promoter	/gene="sgcC"			
		/note="factor Sigma70; predicted +1 start at 4526499"			
		complement(952..982)			
	promoter	/gene="sgcC"			
		/note="factor Sigma70; predicted +1 start at 4526528"			
		complement(960..987)			
	promoter	/gene="sgcC"			
		/note="factor Sigma70; predicted +1 start at 4526536"			
		complement(988..1015)			
	promoter	/gene="sgcC"			
		/note="factor Sigma70; predicted +1 start at 4526564"			
		complement(1046..1075)			
	gene	/gene="sgcC"			
		/note="factor Sigma70; predicted +1 start at 4526622"			
		complement(2515..3666)			
		/gene="sgcX"			
		/note="b4305"			
	CDS	complement(2515..3666)			
		/gene="sgcX"			
		/function="putative enzyme; Not classified"			
		/note="f383; formerly designated yjho"			
		/codon_start=1			
		/transl_table=11			
		/product="putative lyase/synthase"			
		/protein_id="AAC77261.1"			
		/db_xref="GI:2367373"			
		/translation="MIVLRSTEIAMSFSVQETLFSLLQHNNAISGHENAVADVMLCEFR RQAEVWRDRNLGNVARYGSKDPALRMIFAHMDEVGFYVRKITEPSGLFRFERYGGP AQVTMAGSIVTLTGDKGPMGICIGKSYHFAGDERTQSPSVKLMIDIGAKDKDDAI RMIGVGTPTVLYNPPOLLANDLVCSKALDRLGLCTALLGVADAISTWELDIAVLYVA SVQEEFNIRGIVPVLRVVKPDLAIGIDITPSCDTPLDHDYSEVRINQGVGTCLNVHG RGTLAGLITPRLIRMLEQTLAHEHNIPVQREVAFCVITEGYIQVEQGIQFASLSIP CRYTHSPAASLRDLTDCITRLLTALAGMSAAHPFVPPDSTGTTQEAHPL"			
		complement(3819..3848)			
	promoter				


```
/note="factor Sigma70; predicted +1 start at 4529395"
complement(4422..5168)
/feature="yjhp"
/feature="b4306"
complement(4422..5168)
/feature="yjhp"
/function="putative enzyme; Not classified"
/note="f248; 100 pct identical amino acid sequence and
equal length to YJHP_ECOLI SW: P39367"
/codon_start=1
/transl_table=11
/product="putative methyltransferase"
/db_xref="GI:1790760"
/transl_table="MDPRIFTTISEHRIHNPFTTEKYATLGRVLRMKPGTRILDIG
SGSEMLTWARHGITGTIDMSLFTQAQKRAEELGVSERVVHFTHNDAAQVANE
KDVAAQCVATWAGTGAAGAGAEALLAOSLPGGIMLIGEPYWRLOPATTEIAOACGVSS
TDFLTLPLGLVAFDLDGYDVEMVLADQESDWRYEAKWLTHRWLEANPDDDFARE
VRAELNAPRYVYARECGWGVFALIR"
complement(5170..5201)
/note="factor Sigma70; predicted +1 start at 4530746"
complement(5216..5246)
/note="factor Sigma70; predicted +1 start at 4530792"
complement(5224..5769)
/feature="yjhp"
/feature="b4307"
complement(5224..5769)
/feature="yjhp"
/function="orf; Unknown"
/note="f181; 100 pct identical amino acid sequence and
equal length to YJHQ_ECOLI SW: P39368"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAC77263.1"
/db_xref="GI:1790761"
/transl_table="MTVHFTFHITKSDASDIREVETRAFSGKADLVASLLEDES
ARPAISLLARYEGKAVGHILFTFRATKGDMSPLMHILAPLAVIPEYQGMVGGRILR
TGIEHLRLMGCTQFVLGHATYPRHGFEPACDGGKGYADYPIPEEHKACMMQSLTA
QPMTLTGHRICADPDDETALT"
complement(5376..5391)
/feature="yjhp"
/note="central position to predicted promoter: -206"
/bound_moiety="FhlA predicted site"
complement(5376..5391)
/feature="yjhp"
/note="central position to predicted promoter: -160"
/bound_moiety="FhlA predicted site"
complement(5787..5816)
/note="factor Sigma70; predicted +1 start at 4531363"
complement(5806..5835)
/note="factor Sigma70; predicted +1 start at 4531382"
6885..6916
/note="factor Sigma70; predicted +1 start at 4532506"

Query Match 34.3%; Score 26.4; DB 1; Length 12249;
Best Local Similarity 61.8%; Pred. No. 39;
Matches 42; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 10 atgttcaagcgttcgctcactcgggaccggcctaagccgccccttaaccaaa 69
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 9866 ATAGTTCTTACCGTCTCGGTAATTTTCTCGTAACCTTTGAATCCGCCCTCGGCAA 9925
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 70 cggttagag 77
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 9926 TATAAGAG 9933

RESULT 6
AC020814/c
LOCUS
DEFINITION Mus musculus clone RG-MBAC_270F3, LOW-PASS SEQUENCE SAMPLING.
ACCESSION AC020814
```

VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
AUTHORS
TITLE
JOURNAL
COMMENT

AC020814.2 GI:6980191
HTG; HTGS_PHASE0.
house mouse.
Mus musculus
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 28446)
DOE Joint Genome Institute.
Sequencing of Mouse
Unpublished
2 (bases 1 to 28446)
DOE Joint Genome Institute.
Direct Submission
Submitted (10-JAN-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Feb 15, 2000 this sequence version replaced gi:16686494.
* NOTE: This record contains 28 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1 602: contig of 602 bp in length
gap of unknown length
603 1582: contig of 980 bp in length
gap of unknown length
1583 1960: contig of 378 bp in length
gap of unknown length
1961 2981: contig of 1021 bp in length
gap of unknown length
2982 3775: contig of 794 bp in length
gap of unknown length
3776 4866: contig of 1091 bp in length
gap of unknown length
4867 5740: contig of 874 bp in length
gap of unknown length
5741 6748: contig of 1008 bp in length
gap of unknown length
6749 6915: contig of 167 bp in length
gap of unknown length
6916 8285: contig of 1370 bp in length
gap of unknown length
8286 9351: contig of 1066 bp in length
gap of unknown length
9352 10403: contig of 1052 bp in length
gap of unknown length
10404 12170: contig of 1767 bp in length
gap of unknown length
12171 13169: contig of 999 bp in length
gap of unknown length
13170 14130: contig of 961 bp in length
gap of unknown length
14131 14556: contig of 426 bp in length
gap of unknown length
14557 15627: contig of 1071 bp in length
gap of unknown length
15628 16621: contig of 994 bp in length
gap of unknown length
16622 17204: contig of 583 bp in length
gap of unknown length
17205 18280: contig of 1076 bp in length
gap of unknown length
18281 19621: contig of 1341 bp in length
gap of unknown length
19622 20781: contig of 1160 bp in length
gap of unknown length
20782 22156: contig of 1375 bp in length
gap of unknown length

```

* 22157 23397: contig of 1241 bp in length
*      gap of unknown length
* 23398 24986: contig of 1589 bp in length
*      gap of unknown length
* 24987 26187: contig of 1201 bp in length
*      gap of unknown length
* 26188 27056: contig of 869 bp in length
*      gap of unknown length
* 27057 28446: contig of 1390 bp in length.
      Location/Qualifiers
        ..28446
        /organism="Mus musculus"
        /db_xref="taxon:10090"
        /clone="RG-MBAC-270F3"
BASE COUNT 7571 a 7154 c 6705 g 6946 t 70 others
ORIGIN

Query Match 34.3%; Score 26.4; DB 66; Length 28446;
Best Local Similarity 61.8%; Pred. No. 38;
Matches 42; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 10 atggttcagcgttcgttcctcactcgagaccggttaagccggcccttaaccacaa 69
   || || || || || || || || || || || || || || || || || || || ||
Db 16694 ATAGTCTTACCGTCTCGTAATTTCTCGTAACCTTTGAATCGGCCCTCCGGCAA 16635

QY 70 cuttagag 77
   | ||||
Db 16634 TATAAGAG 16627

RESULT 7
AC011073/c
LOCUS Homo sapiens chromosome 17 clone RP11-51E18 map 17, WORKING DRAFT
DEFINITION AC011073
ACCESSION AC011073.3 GI:7341685
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS human.
SOURCE Homo sapiens
  ORGANISM Homo sapiens
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  1 (bases 1 to 160312)
    Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
    Baldwin,J., Barna,N., Beckerly,R., Bozulavsky,L., Bouckgalter,B.,
    Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
    Cooke,P., Dearellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
    Ferrelira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
    Galagan,J., Gardyna,S., Grant,G., Hagos,B., Hearford,A., Horton,L.,
    Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
    Lehotsky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
    McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
    Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
    Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
    Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
    Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
    Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
  Direct Submission
  Submitted (01-OCT-1999) Whitehead Institute/MIT Center for Genome
  Research, 320 Charles Street, Cambridge, MA 02141, USA
  On Mar 30, 2000 this sequence version replaced gi:6042048.
  All repeats were identified using RepeatMasker:
  Smit, A.F.A. & Green, P. (1996-1997)
  http://ftp.genome.washington.edu/RM/RepeatMasker.html
  -----
  Center: Whitehead Institute/ MIT Center for Genome Research
  Center code: WIBR
  Web site: http://www-seq.wi.mit.edu

```

```

Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L1039
Center clone name: 51_E_18
----- Summary Statistics
Sequencing vector: M13; M7815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 116443 bases at least Q40
Consensus quality: 137211 bases at least Q30
Consensus quality: 151029 bases at least Q20
Insert size: 165000; agarose-fp
Quality coverage: 3.4 in Q20 bases; agarose-fp
Quality coverage: 3.5 in Q20 bases; sum-of-contigs
-----

```

```

* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.

```

```

1 3935: contig of 3935 bp in length
3936 4035: gap of 100 bp
4036 8766: contig of 4731 bp in length
8767 8866: gap of 100 bp
8867 16083: contig of 7217 bp in length
16084 16183: gap of 100 bp
16184 41851: contig of 25668 bp in length
41852 41951: gap of 100 bp
41952 69493: contig of 27542 bp in length
69494 69593: gap of 100 bp
69594 160312: contig of 90719 bp in length.

```

FEATURES

source

```

1.160312
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="17"
/map="17"
/clone="RP11-51E18"
/clone_lib="RPC1-11 Human Male BAC"
1..3935

```

misc_feature

```

/note="assembly_fragment"
4036..8766

```

misc_feature

```

/note="assembly_fragment"
clone_end:SP6
vector_side:left
8867..16083

```

misc_feature

```

/note="assembly_fragment"
16184..41851

```

misc_feature

```

/note="assembly_fragment"
41952..69493

```

misc_feature

```

/note="assembly_fragment"
69594..160312

```

misc_feature

```

/note="assembly_fragment"
BASE COUNT 47017 a 32095 c 32748 g 47949 t
ORIGIN

```

Query Match

34.3%; Score 26.4; DB 62; Length 160312;

Best Local Similarity

69.2%; Pred. No. 35;

Matches

36; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY

23 ttgcgttcgttcactcgagaccggttaagccggcccttaaccacgtta 74

|| || || || || || || || || || || || || || || || || || || ||

Db 23733 TTGCTTTTCCTCACTTTTGATTTGACTAAAGCTGGCCACTTGGCCCAACCTAA 23682

RESULT

8

AC006461

LOCUS

AC006461 181215 bp DNA PRI 21-DEC-1999

```

DEFINITION Homo sapiens BAC clone RP11-343N14 from 2, complete sequence.
ACCESSION AC006461
VERSION AC006461.2 GI:4508128
KEYWORDS HTG.
SOURCE human.
ORGANISM
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
  1 (bases 1 to 181215)
  Sulston, J.E. and Waterston, R.
  Toward a complete human genome sequence
  Genome Res. 8 (11), 1097-1108 (1998)
99063792
2 (bases 1 to 181215)
  Kalicki, J., Bauer, C. and Smith, R.
  The sequence of Homo sapiens BAC clone RP11-343N14
  Unpublished
3 (bases 1 to 181215)
  Waterston, R.H.
  Direct Submission
  Submitted (28-JAN-1999) Genome Sequencing Center, Washington
  University School of Medicine, 4444 Forest Park Parkway, St. Louis,
  MO 63108, USA
4 (bases 1 to 181215)
  Waterston, R.H.
  Direct Submission
  Submitted (24-MAR-1999) Genome Sequencing Center, Washington
  University School of Medicine, 4444 Forest Park Parkway, St. Louis,
  MO 63108, USA
5 (bases 1 to 181215)
  Waterston, R.
  Direct Submission
  Submitted (22-OCT-1999) Department of Genetics, Washington
  University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
6 (bases 1 to 181215)
  Waterston, R.
  Direct Submission
  Submitted (21-DEC-1999) Department of Genetics, Washington
  University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
  On Mar 24, 1999 this sequence version replaced gi:4199979.
----- Summary Statistics
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@wustl.wustl.edu
-----
Center project name: H_NH0343N14
-----
NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. John D.
McPherson, Department of Genetics, Washington University, St. Louis
MO. For additional information about the map position of this
sequence, see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
The RPCI-11 human BAC library was made from the blood of one male
donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E.,
Tateno, M., Catanese, J.J. and de Jong, P.J. (1998) An improved
approach for construction of bacterial artificial chromosome
libraries. Genomics 51:1-8. The clone may be obtained either from
Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
and coworkers at the Roswell Park Cancer Institute
(http://bacpac.med.buffalo.edu)
VECTOR: pBACe3.6
NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the right is RP11-493L16. Actual start of
this clone is at base position 1 of RP11-343N14; actual end is at
base position 181215 of RP11-343N14.
Location/Qualifiers
  1. 181215
  /organism="Homo sapiens"
  /db_xref="taxon:9606"
  /chromosome="2"
  /map="2"
  /clone="RP11-343N14"
  /clone_lib="RPCI-11"
  117. 414
  /rpt_family="MERL_type"
  repeat_region
  1915. 2634
  /rpt_family="L1"
  repeat_region
  2635. 4474
  /rpt_family="L1"
  repeat_region
  4475. 4777
  /rpt_family="Alu"
  repeat_region
  4778. 4842
  /rpt_family="L1"
  repeat_region
  4880. 5125
  /rpt_family="L1"
  repeat_region
  5126. 5150
  /rpt_family="(TTG)n"
  repeat_region
  5154. 5268
  /rpt_family="Alu"
  repeat_region
  5269. 5556
  /rpt_family="L1"
  repeat_region
  5554. 6548
  /rpt_family="L1"
  repeat_region
  6543. 6609
  /rpt_family="L1"
  repeat_region
  6606. 6709
  /rpt_family="L1"
  repeat_region
  7833. 7851
  /note="similar to EST AA923166 (NID:g3070475) on18h05.s1"
  misc_feature
  7939. 8000
  /rpt_family="AT-rich"
  repeat_region
  8936. 9014
  /rpt_family="(TG)n"
  repeat_region
  9329. 9357
  /rpt_family="AT-rich"
  repeat_region
  9646. 9864
  /rpt_family="MIR"
  repeat_region
  10892. 11196
  /rpt_family="Alu"
  repeat_region
  10901. 10910
  /note="similar to EST A1288780 (NID:g3932283) qm06f04.x1"
  repeat_region
  11878. 11901
  /rpt_family="(TG)n"
  repeat_region
  12323. 12372
  /rpt_family="(TG)n"
  repeat_region
  gene
  <12903. 168586
  /gene="WUGSC:H_NH0343N14.1"
  CDS
  join(<12903. 12974, 163734. 163846, 168460. 168586)
  /gene="WUGSC:H_NH0343N14.1"
  /note="similar to rsec15 [Rattus norvegicus]; similar to
  AAC01580.1 (PID:g2B27162); H_NH0343N14.1"
  /codon_start=1
  /evidence=not_experimental
  /protein_id="AAF03520.1"
  /db_xref="GI:6094668"
  /translation="FARSGPVGCFQEDTLQALFIDLRLDLFIQDWSTYLADYGOP
  S"
  misc_feature
  13835. 13846

```



```

KGSVSCGIVSDVNDIDIVRLMVGRELONRNNAMKENVSNLAHETVFEVRNVTSDR
KYRDISVCGEILGAGLVGSGRTLMNCLFGVDKRAGGEIRUNGKDISPRSLD
AYKKMAYITESRGRTFPNFSIAQNNMISRLCKDGKYGKAGFLFHEVDEQTAENQ
RELLAKKHSVNNQITELSGNQKVLISKWLCCCPVEIIFDEPTRGIDVGAKEIYK
VNRQLADGCKVILMVSSSELPEIITVCDRIAVFCESGLTQILTNRDDMSEETIMAWALP
OR".
misc_structure      809.. 931
/feature="predicted bend of 74.75 degrees"
/feature="predicted bend of 75.87 degrees"
/feature="predicted bend of 75.87 degrees"
/feature="promoter matrix score of 55; putative; resembles
flh-regulated promoter"
/standard_name="REP; repetitive extragenic palindromic
element"
/feature="contains 1 REP sequence"
complement(2036..2971)
/feature="ORF_f311"
/codon_start=1
/transl_table=11
/label="ORF_f311"
/protein_id="AAA96987.1"
/db_xref="GI:536932"
/translation="MNKYLKYSFSGTLVGLMLSTSFAAAEYAVVLKTLNPNFVDMKK
GIEDAKTLGVSVDIFAPSGSEDFQSOLQFEDLSNKNYKGIAFAPLSVNLVMPYAR
AWKGIYLVNDEKIDMNLKAGNVEAFVTTDNVADVGAKGASFIIDKLGAEGEVA
IIEKGAGNAGSARNGATEAPFKASQIKLVASQPADWDRIKALDVATNVLORNPIK
ATYCANDTMAGVACAVANAGKTGKLVVGTGDIPEARWVEAGQATATVAQNPDIG
ATGLKLVDAEKSQVPLDKAPEKLVDSILVTQ"
misc_structure      3010..3121
/feature="predicted bend of 73.22 degrees"
complement(3030..3953)
/feature="ORF_f307"
/codon_start=1
/transl_table=11
/label="ORF_f307"
/protein_id="AAA96988.1"
/db_xref="GI:536933"
/translation="WASSIQOENVMWSQFSDALPNIGLAPYLRMKOEGMTENESR
IVELLLKQNLSCAPAIKDVAALAVSEAMIVKSLKLGFGFRNLRSALDYFSQSE
QVLPSELAFDEAPQDVNVKNFNTIRTHEGOSIVNDEIHRRAARFFQARORDLYGA
GNSNACADVQHKRIGRVCQAYPDHIMMSASLQEGDVLVVYVTHSGRTSDYKAA
VELAKNGAKICITHSYHSPITAKLADYICSPAPETPLLRNASARILQLTLLDAFF
VSAQLNTEQANINMOKTCAIYDFSPGALK"
promoter            3952..3979
/feature="promoter matrix score of 61; putative"
misc_structure      4151..4251
/feature="predicted bend of 72.06 degrees"
4279..4728
/feature="y1ca"
4279..4728
/feature="y1ca"
/codon_start=1
/transl_table=11
/label="ORF_0149"
/protein_id="AAA96989.1"
/db_xref="GI:536934"
/translation="MKIATGCDHVGFILKHEIVAHLYVERGVEVIDKGTWSSERTDYP
HYASQVALAVAGGVEDGILLICGTGVSITANKFAGIRAVVCSEPYSAQLSRQHNDT
NVLAFGRVGLGELAKMIVDAWLGAQYEGRRHQQRVEAITAEQRN"
misc_feature        4520..16191
/feature="corresponds to GenBank Accession Number D90227
(ECOPHN) 1..11672; the related sequence GenBank Accession
Number J05260 (ECOPHNAQ) is from E. coli B, not K-12"
complement(4739..5146)
/feature="phnQ"
complement(4739..5146)
/feature="phnQ"
/feature="annotated in GenBank Accession Number J05260
(ECOPHNAQ), but probably not a gene"
/codon_start=1
/transl_table=11

```

```

/protein_id="AAA96990.1"
/db_xref="GI:536935"
/translation="MPAPYTSPIILIFHSPLFIQLPTLLIKLALAIVOLPAAIKFL
AAAVILSAPIVAHFTLPVAVARIFIVSPTRVISVIRNLLNIANGVVDNIIGLRNVD
SVSRVKNRSGKRFISQPKETLSTLIITHPC"
complement(5272..6030)
/feature="phnP"
complement(5272..6030)
/feature="phnP"
/codon_start=1
/transl_table=11
/protein_id="AAA96991.1"
/db_xref="GI:536936"
/translation="MSLTITLTGTGAQGVPAWGCECAACARARRSPQYRPOCPSGVV
KENDATLIDAGLHDLADRWSPGSGFOQLLTHYMDHVOCGLFPLRWGCGVPIPVYGP
DSQGDLDLFEKPGCLLDFSHTEVFVFDLQGVTLPLLNHSLKTLFGYLLLETAHSRVA
WLSDTAGLFEKTLKFLRNQOPQVMVDCSHPPRADAPRNHCDLNTVLALNQVITRSPRV
ILTHISHQFDAMLNENALPSGFVEGDFGMEIGVA"
misc_structure      5452..5586
/feature="predicted bend of 75.60 degrees"
complement(6032..6466)
/feature="phnO"
complement(6032..6466)
/feature="phnO"
/codon_start=1
/transl_table=11
/protein_id="AAA96992.1"
/db_xref="GI:536937"
/translation="MPACELRPATQYTDVAYALICELKQAEFDHIAFRVGFNANLRD
PNMRYHALLDGEVVMIGLHLQFHLHVNMIIGEIVVMPOARGLNVGSKLLAWAE
EPARQAGAEMTSLTNVNRHDAHRFYLRGYESQSHFRFTKAL"
misc_structure      6453..7010
/feature="phnN"
complement(6453..7010)
/feature="phnN"
/feature="HisP-like nucleotide binding protein (ATP-binding
protein)"
/codon_start=1
/transl_table=11
/protein_id="AAA96993.1"
/db_xref="GI:536938"
/translation="MNGKLIWLMGSPSGKDSLAEILRLREOTQLLVAHRYITRDASA
GSENHIALSEQEFFTRAGONLLALSHWANGLYGVGVEIDLWLHAGFDVLVNGSRHL
POARARYQSALLPVCLQVSPETILRORLENRGRENASEINARLARARYTPQDCHTLNN
DGSLSRVSFTLLTLIHQEKHACL"
misc_structure      7010..8146
/feature="phnM"
complement(7010..8146)
/feature="phnM"
/feature="HisM-like integral membrane protein"
Query Match      34.3%; Score 26.4; DB 2; Length 338534;
Best Local Similarity 61.8%; Pred. No. 33;
Matches 42; Conservative 0; Mismatches 26; Indels 0; Caps 0;
QY 10 atgttcaagcttcgcttcgctcactcgggaccggcgttaagccggcccttaaccaa 69
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 228713 ATAGTTCTTACCGTTCGGTAATTTTCTCGTGAACCTTTGAATCCGCGCCCTCGGCAAA 228772
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 70 cgttagag 77
| | | | |
Db 228773 TATAACAG 228780
| | | | |
RESULT 10
AL353773/c
LOCUS
DEFINITION
Homo sapiens chromosome 9 clone RP11-574M5, *** SEQUENCING IN
PROGRESS ***, 3 unordered pieces.
ACCESSION
AL353773
VERSION
AL353773.6 GI:13161600
KEYWORDS
HTG; HTGS_PHASE1; HTGS_DRAFT.
human.

```

```

ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 181931)
Direct Submission
Submitted (24-FEB-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Feb 28, 2001 this sequence version replaced gi:10129433.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA574M5
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 180786 bases at least Q40
Consensus quality: 181295 bases at least Q30
Consensus quality: 181617 bases at least Q20
Insert size: 181731; sum-of-contigs
Insert size: 187130; 5.8% error; agarose-fp
Quality coverage: 6.60x in Q20 bases; sum-of-contigs Quality
coverage: 6.75x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 7055: contig of 7055 bp in length
* 7056 7155: gap of 100 bp
* 7156 43164: contig of 36009 bp in length
* 43165 43264: gap of 100 bp
* 43265 181931: contig of 138667 bp in length.
FEATURES             Location/Qualifiers
     source           1..181931
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="9"
                     /clone_lib="RP11-574M5"
                     /clone="RP11-574M5"
                     /clone="RP11-574M5"
misc_feature         1..7055
                     /note="assembly_fragment:02849"
misc_feature         7156..43164
                     /note="assembly_fragment:02176"
                     /fragment_chain:1
misc_feature         43265..181931
                     /note="assembly_fragment:02673"
                     /fragment_chain:1
                     clone_end:SP6
                     vector_side:right
BASE COUNT          52824 a 35727 c 36646 g 56533 t 201 others
ORIGIN
Query Match          33.8%; Score 26; DB 79; Length 181931;
Best Local Similarity 59.5%; Pred. No. 48;
Matches 44; Conservative 0; Mismatches 30; Indels 0; Gaps 0;
QY      3 ctaacaatggttcaaacgcttcgcttcactcgaggccggaacccgcccctt 62
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 108165 CCACCAACTGGTAGCTCCCTCCCTTCCTCATTCAGTACCACACATATCTATACCCCT 108106
QY      63 aaccaaacgttaga 76
        ||| ||| ||| |||

```

```

Db 108105 CTCCAATAATAGA 108092
RESULT 11
AC006239/c
LOCUS      AC006239 195032 bp DNA PRI 31-MAR-1999
DEFINITION Homo sapiens chromosome 9, clone hRPK.467_F_21, complete sequence.
AC006239
ACCESSION AC006239.5 GI:4544480
VERSION
KEYWORDS   human.
SOURCE     Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
            1 (bases 1 to 195032)
            Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,
            Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., Dearellano,K., Depayre,E., Devon,K., Dewar,K.,
            Donelan,L., Ferreira,P., Fitzhugh,W., Forrest,C., Funke,R.,
            Gage,D., Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A.,
            Herena,L., Horton,L., Howland,J.C., Jones,C., Kann,L., Karatas,A.,
            Lehoczy,J., Macdonald,P., Marquis,N., McEwan,P., McGurk,A.,
            McKernan,K., Meldrim,J., Molla,M., Morris,W., Morrow,J.,
            Mychaleckyj,J., Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P.,
            Pavlin,B., Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A.,
            Severy,P., Stange-Thomann,N., Stojanovic,N., Stone,C.,
            Subramanian,A., Tesfaye,S., Torruella-Miller,I., Vassiliev,H.,
            Vo,A., Wagner,A., Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
            Direct Submission
            Submitted (18-DEC-1998) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
            3 (bases 1 to 195032)
            Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,
            Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., Dearellano,K., Depayre,E., Devon,K., Dewar,K.,
            Donelan,L., Doyle,M., Ferreira,P., Fitzhugh,W., Forrest,C.,
            Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
            Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
            Karatas,A., Lehoczy,J., Lieu,C., Locke,K., Macdonald,P.,
            Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
            Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
            Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
            Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
            Tesfaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
            Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
            Direct Submission
            Submitted (31-MAR-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
            On Mar 30, 1999 this sequence version replaced gi:4529984.
            All repeats were identified using RepeatMasker: Smit, A.F.A. &
            Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html.
FEATURES             Location/Qualifiers
     source           1..195032
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="hRPK.467_F_21"
                     /clone_lib="RP11-11 human BAC library"
                     /map="9"
                     /chromosome="9"
                     complement(421..576)
                     /rpt_family="MIR"
                     819..911
                     /rpt_family="L2"
                     1174..1297
                     /rpt_family="MIR"
                     repeat_region
                     repeat_region
                     repeat_region

```

```
repeat_region complement(1571..1724)
repeat_region /rpt_family="MER5A"
repeat_region complement(2457..2525)
repeat_region /rpt_family="MIR"
repeat_region 2608..2639
repeat_region /rpt_family="(TTCA)n"
repeat_region complement(2640..2779)
repeat_region /rpt_family="L2"
repeat_region complement(2963..3413)
repeat_region /rpt_family="MLT1E2"
repeat_region complement(3634..4138)
repeat_region /rpt_family="MLT1H"
repeat_region complement(4232..4698)
repeat_region /rpt_family="LJMA9"
repeat_region complement(4699..5019)
repeat_region /rpt_family="MER2"
repeat_region complement(5020..9117)
repeat_region /rpt_family="LJMA9"
repeat_region complement(9118..9428)
repeat_region /rpt_family="AluJo"
repeat_region complement(9429..9603)
repeat_region /rpt_family="LJMA9"
repeat_region complement(9603..10488)
repeat_region /rpt_family="LJM2"
repeat_region complement(11949..12159)
repeat_region /rpt_family="LJM2"
repeat_region complement(12306..12420)
repeat_region /rpt_family="LJM2"
repeat_region 12586..12636
repeat_region /rpt_family="(TTG)n"
repeat_region complement(12735..13104)
repeat_region /rpt_family="HERVL"
repeat_region 13107..13188
repeat_region /rpt_family="MIR"
repeat_region 14010..14091
repeat_region /rpt_family="MIR"
repeat_region complement(14273..15117)
repeat_region /rpt_family="LJME3A"
repeat_region complement(15118..15474)
repeat_region /rpt_family="MLT1A1"
repeat_region complement(15475..15836)
repeat_region /rpt_family="LJME3A"
repeat_region complement(15930..16099)
repeat_region /rpt_family="MIR"
repeat_region complement(16421..16490)
repeat_region /rpt_family="MIR"
repeat_region 17393..17423
repeat_region /rpt_family="(CA)n"
repeat_region 18109..18327
repeat_region /rpt_family="MIR"
repeat_region complement(20765..20839)
repeat_region /rpt_family="L2"
repeat_region complement(21217..21364)
repeat_region /rpt_family="MIR"
repeat_region 21588..21619
repeat_region /rpt_family="(CACCAT)n"
repeat_region complement(21793..21958)
repeat_region /rpt_family="MIR"
repeat_region 22369..22374
repeat_region /rpt_family="(T)n"
repeat_region complement(22375..22753)
repeat_region /rpt_family="LJPA7"
repeat_region 22754..22774
repeat_region /rpt_family="(T)n"
repeat_region complement(22775..28038)
repeat_region /rpt_family="LJPA2"
repeat_region complement(28034..28817)
repeat_region /rpt_family="LJPA2"
repeat_region complement(28815..29104)
repeat_region /rpt_family="LJP1"
repeat_region complement(29461..29531)
repeat_region /rpt_family="MIR"
repeat_region 30893..30928

/rpt_family="(CAAAA)n"
31786..31806
/rpt_family="AT_rich"
complement(33223..33412)
/rpt_family="L2"
34813..34995
/rpt_family="LJM4"
35075..35115
/rpt_family="AT_rich"
35298..35322
/rpt_family="(CA)n"
36949..36970
/rpt_family="AT_rich"
37155..37374
/rpt_family="MIR"
37346..37396
/rpt_family="L2"
complement(37450..37530)
/rpt_family="MIR"
complement(37531..37897)
/rpt_family="MLT1A1"
complement(37898..38035)
/rpt_family="MIR"
complement(39084..39196)
/rpt_family="MIR"
39250..39517
/rpt_family="MIR"
39774..40063
/rpt_family="AluSg"
complement(40557..40909)
/rpt_family="THE1A"
complement(41349..41511)
/rpt_family="L2"
42064..42309
/rpt_family="MIR"
42422..42476
/rpt_family="AT_rich"
42699..42732
/rpt_family="(CAA)n"
42908..43312
/rpt_family="MLT1B"
43916..44036
/rpt_family="L2"
complement(44771..44884)
/rpt_family="MIR"
complement(44967..45018)
/rpt_family="L2"
45218..45804
/rpt_family="MER77"
complement(46858..46962)
/rpt_family="L2"
complement(48753..48876)
/rpt_family="MIR"
49245..49295
/rpt_family="AT_rich"
complement(49599..49899)
/rpt_family="MLT1J2"
50108..50158
/rpt_family="(CATTT)n"

Query Match 33.8%; Score 26; DB 86; Length 195032;
Best Local Similarity 59.5%; Pred. No. 48;
Matches 44; Conservative 0; Mismatches 30; Indels 0; Gaps 0;
```

```
Oy 3 ctacaacaaagtgcacgcgttcgttcgctcactcggaccggtacaaagccggccctt 62
Db 12755 CCACCAACTGGTAGCATCCCTCCCTTCCTCATTCCAGTACCCACACTATCTATACCCCT 12696
Oy 63 aaccaaagcttaga 76
Db 12695 CTCCAATAATAGA 12682
```


[illegible]

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 2, 2001, 16:53:10 ; Search time 183.37 Seconds
(without alignments)
263.666 Million cell updates/sec

Title: US-09-701-626A-78
Perfect score: 77
Sequence: 1 cttcaacaatggttcaagc.....cccttaacacaaacgttagag 77

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_0601.*
1: /cgnl_9/gcgdata/geneseq/geneseq/NA1980.DAT.*
2: /cgnl_9/gcgdata/geneseq/geneseq/NA1981.DAT.*
3: /cgnl_9/gcgdata/geneseq/geneseq/NA1982.DAT.*
4: /cgnl_9/gcgdata/geneseq/geneseq/NA1983.DAT.*
5: /cgnl_9/gcgdata/geneseq/geneseq/NA1984.DAT.*
6: /cgnl_9/gcgdata/geneseq/geneseq/NA1985.DAT.*
7: /cgnl_9/gcgdata/geneseq/geneseq/NA1986.DAT.*
8: /cgnl_9/gcgdata/geneseq/geneseq/NA1987.DAT.*
9: /cgnl_9/gcgdata/geneseq/geneseq/NA1988.DAT.*
10: /cgnl_9/gcgdata/geneseq/geneseq/NA1989.DAT.*
11: /cgnl_9/gcgdata/geneseq/geneseq/NA1990.DAT.*
12: /cgnl_9/gcgdata/geneseq/geneseq/NA1991.DAT.*
13: /cgnl_9/gcgdata/geneseq/geneseq/NA1992.DAT.*
14: /cgnl_9/gcgdata/geneseq/geneseq/NA1993.DAT.*
15: /cgnl_9/gcgdata/geneseq/geneseq/NA1994.DAT.*
16: /cgnl_9/gcgdata/geneseq/geneseq/NA1995.DAT.*
17: /cgnl_9/gcgdata/geneseq/geneseq/NA1996.DAT.*
18: /cgnl_9/gcgdata/geneseq/geneseq/NA1997.DAT.*
19: /cgnl_9/gcgdata/geneseq/geneseq/NA1998.DAT.*
20: /cgnl_9/gcgdata/geneseq/geneseq/NA1999.DAT.*
21: /cgnl_9/gcgdata/3'', seq/geneseq/NA2000.DAT.*
22: /cgnl_9/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	77	100.0	77	21	AAZ44967 P. alcaligenes rep
2	74	96.1	74	21	AAZ88516 P. alcaligenes rep
3	74	96.1	77	21	AAZ44966 P. alcaligenes rep
4	74	96.1	14143	21	AAZ44981 P. alcaligenes res
5	71.8	93.2	77	21	AAZ44908 P. alcaligenes rep
6	71.4	92.7	74	21	AAZ88511 P. alcaligenes rep
7	71.4	92.7	77	21	AAZ44952 P. alcaligenes rep
8	71.4	92.7	77	21	AAZ44987 P. alcaligenes rep
9	71.2	92.5	77	21	AAZ44954 P. alcaligenes rep
10	69.8	90.6	77	21	AAZ44900 P. alcaligenes rep
11	69.8	90.6	77	21	AAZ44913 P. alcaligenes rep

12	69.2	89.9	77	21	AAZ44944 P. alcaligenes rep
13	68.6	89.1	77	21	AAZ44911 P. alcaligenes rep
14	68.2	88.6	74	21	AAZ88505 P. alcaligenes rep
15	68.2	88.6	74	21	AAZ88515 P. alcaligenes rep
16	68.2	88.6	74	21	AAZ88520 P. alcaligenes rep
17	68.2	88.6	77	21	AAZ44930 P. alcaligenes rep
18	68.2	88.6	77	21	AAZ44957 P. alcaligenes rep
19	68.2	88.6	13382	21	AAZ44984 P. alcaligenes res
20	67.6	87.8	77	21	AAZ44984 P. alcaligenes rep
21	67.4	87.5	77	21	AAZ44937 P. alcaligenes rep
22	67.4	87.5	7300	21	AAZ44982 P. alcaligenes res
23	67.4	87.5	7300	21	AAZ44983 P. alcaligenes res
24	67	87.0	77	21	AAZ44910 P. alcaligenes rep
25	67	87.0	77	21	AAZ44990 P. alcaligenes rep
26	66.6	86.5	74	21	AAZ88503 P. alcaligenes rep
27	65.4	84.9	77	21	AAZ44964 P. alcaligenes rep
28	65	84.4	77	21	AAZ44915 P. alcaligenes rep
29	65	84.4	77	21	AAZ44929 P. alcaligenes rep
30	64.4	83.6	78	21	AAZ44907 P. alcaligenes rep
31	62.4	81.0	73	21	AAZ88521 P. alcaligenes rep
32	62.4	81.0	78	21	AAZ44933 P. alcaligenes rep
33	61.2	79.5	78	21	AAZ44936 P. alcaligenes rep
34	60.8	79.0	78	21	AAZ44938 P. alcaligenes rep
35	60.4	78.4	78	21	AAZ44912 P. alcaligenes rep
36	60	77.9	78	21	AAZ44963 P. alcaligenes rep
37	59.2	76.9	78	21	AAZ44940 P. alcaligenes rep
38	58.2	75.6	78	21	AAZ44927 P. alcaligenes rep
39	57.2	74.3	75	21	AAZ88508 P. alcaligenes rep
40	57.2	74.3	76	21	AAZ44961 P. alcaligenes rep
41	57.2	74.3	78	21	AAZ44997 P. alcaligenes rep
42	56.2	73.0	78	21	AAZ44948 P. alcaligenes rep
43	56	72.7	78	21	AAZ44956 P. alcaligenes rep
44	55.8	72.5	76	21	AAZ44962 P. alcaligenes rep
45	55.8	72.5	78	21	AAZ44928 P. alcaligenes rep

ALIGNMENTS

RESULT 1
AAZ44967
ID AAZ44967 standard; DNA; 77 BP.
XX
AC AAZ44967;
XX
DT 16-MAY-2000 (first entry)
XX
DE P. alcaligenes repeat (PAR) element DNA #74.
XX
KW Diversity-selected gene; restriction enzyme; adhesin; toxin;
KW detoxifying enzyme; repeat element; PAR; ss.
XX
OS Pseudomonas alcaligenes.
XX
PN WO9964632-Al.
XX
PD 16-DEC-1999.
XX
PF 11-JUN-1999; 99WO-US13295.
XX
PR 12-JUN-1998; 98US-0089086.
PR 12-JUN-1998; 98US-0089101.
XX
PA (NEWE) NEW ENGLAND BIOLABS INC.
XX
PI Raleigh EA, Vaisvila R, Morgan RD;
XX
DR WPI; 2000-116558/10.
XX
PT Cloning intact genes used to isolate genes for restriction enzymes
XX
PS Claim 7a; Page 61; 97pp; English.
XX

CC This invention describes a novel method for cloning intact,
 CC diversity-selected genes (I) from within gene cassettes (GC) which
 CC comprises identifying DNA repeats that flank GC, hybridizing
 CC oligonucleotides (ON) to these repeats and amplification to produce DNA
 CC fragments containing (I), ligating these fragments into a vector and
 CC transforming cells with the vector. This method is used to clone a wide
 CC variety of prokaryotic genes that provide a selective advantage under
 CC particular conditions, particularly those that encode restriction
 CC enzymes (used as reagents in molecular biology); adhesins (for use in
 CC coating or for targeting molecules or organisms to particular sites,
 CC e.g. for competitive exclusion of a selected pathogen); detoxifying
 CC inhibitors or antagonists of the toxin, or in vaccination, or a
 CC modification methyltransferase. Intact genes can be cloned directly with
 CC a high probability that the orientation of expression is known in advance
 CC and low probability of association with extraneous, possibly toxic,
 CC genes. AAZ4894-244980 represent the *Pseudomonas* alcaligenes repeat (PAR)
 CC elements described in the method of the invention.
 XX
 SQ Sequence 77 BP; 19 A; 25 C; 17 G; 16 T; 0 other;

Query Match 100.0%; Score 77; DB 21; Length 77;
 Best Local Similarity 100.0%; Pred. No. 8.4e-20;
 Matches 77; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 ctctacaacaaatggttcaagcgttcgttcgtcactcgcggacggttaagcggcccc 60
 Db 1 ctctacaacaaatggttcaagcgttcgttcgtcactcgcggacggttaagcggcccc 60
 QY 61 ttaacaaacgttagag 77
 Db 61 ttaacaaacgttagag 77

RESULT 2
 AAZ88516
 ID AAZ88516 standard; DNA; 74 BP.
 XX
 AC AAZ88516;
 DT 16-MAY-2000 (first entry)
 DE
 XX P. alcaligenes repeat (PAR) element DNA PARf13.
 XX Diversity-selected gene; restriction enzyme; adhesin; toxin;
 KW detoxifying enzyme; repeat element; PAR; ss.
 XX
 OS *Pseudomonas* alcaligenes.
 XX
 PN WO9964632-A1.
 XX
 PD 16-DEC-1999.
 XX
 PF 11-JUN-1999; 99WO-US13295.
 XX
 PR 12-JUN-1998; 98US-0089086.
 PR 12-JUN-1998; 98US-0089101.
 XX
 PA (NEW) NEW ENGLAND BIOLABS INC.
 XX
 PI Raleigh EA, Vaisvila R, Morgan RD;
 XX
 DR WPI; 2000-116558/10.

CC Cloning intact genes used to isolate genes for restriction enzymes -
 XX
 XX Example 1B; Fig 3E; 97pp; English.
 XX
 CC This invention describes a novel method for cloning intact,
 CC diversity-selected genes (I) from within gene cassettes (GC) which
 CC comprises identifying DNA repeats that flank GC, hybridizing
 CC oligonucleotides (ON) to these repeats and amplification to produce DNA

CC fragments containing (I), ligating these fragments into a vector and
 CC transforming cells with the vector. This method is used to clone a wide
 CC variety of prokaryotic genes that provide a selective advantage under
 CC particular conditions, particularly those that encode restriction
 CC enzymes (used as reagents in molecular biology); adhesins (for use in
 CC coating or for targeting molecules or organisms to particular sites,
 CC e.g. for competitive exclusion of a selected pathogen); detoxifying
 CC enzymes; toxins that interact with a host, e.g. for synthesis of
 CC inhibitors or antagonists of the toxin, or in vaccination, or a
 CC modification methyltransferase. Intact genes can be cloned directly with
 CC a high probability that the orientation of expression is known in advance
 CC and low probability of association with extraneous, possibly toxic,
 CC genes. AAZ89504-288521 represent *Pseudomonas* alcaligenes repeat (PAR)
 CC elements described in the method of the invention.
 XX
 SQ Sequence 74 BP; 19 A; 24 C; 16 G; 15 T; 0 other;

Query Match 96.1%; Score 74; DB 21; Length 74;
 Best Local Similarity 100.0%; Pred. No. 1.1e-18;
 Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 3 ctacaacaaatggttcaagcgttcgttcgtcactcgcggacggttaagcggcccc 62
 Db 1 ctacaacaaatggttcaagcgttcgttcgtcactcgcggacggttaagcggcccc 60
 QY 63 aacaaacgttaga 76
 Db 61 aacaaacgttaga 74

RESULT 3
 AAZ44966
 ID AAZ44966 standard; DNA; 77 BP.
 XX
 AC AAZ44966;
 DT 16-MAY-2000 (first entry)
 DE
 XX P. alcaligenes repeat (PAR) element DNA #73.
 XX Diversity-selected gene; restriction enzyme; adhesin; toxin;
 KW detoxifying enzyme; repeat element; PAR; ss.
 XX
 OS *Pseudomonas* alcaligenes.
 XX
 PN WO9964632-A1.
 XX
 PD 16-DEC-1999.
 XX
 PF 11-JUN-1999; 99WO-US13295.
 XX
 PR 12-JUN-1998; 98US-0089086.
 PR 12-JUN-1998; 98US-0089101.
 XX
 PA (NEW) NEW ENGLAND BIOLABS INC.
 XX
 PI Raleigh EA, Vaisvila R, Morgan RD;
 XX
 DR WPI; 2000-116558/10.

CC Cloning intact genes used to isolate genes for restriction enzymes -
 XX
 XX Claim 7a; Page 61; 97pp; English.
 CC
 CC This invention describes a novel method for cloning intact,
 CC diversity-selected genes (I) from within gene cassettes (GC) which
 CC comprises identifying DNA repeats that flank GC, hybridizing
 CC oligonucleotides (ON) to these repeats and amplification to produce DNA
 CC fragments containing (I), ligating these fragments into a vector and
 CC transforming cells with the vector. This method is used to clone a wide
 CC variety of prokaryotic genes that provide a selective advantage under
 CC particular conditions, particularly those that encode restriction

CC enzymes (used as reagents in molecular biology); adhesins (for use in
CC coating or for targeting molecules or organisms to particular sites,
CC e.g. for competitive exclusion of a selected pathogen); detoxifying
CC enzymes; toxins that interact with a host, e.g. for synthesis of
CC inhibitors or antagonists of the toxin, or in vaccination, or a
CC modification methyltransferase. Intact genes can be cloned directly with
CC a high probability that the orientation of expression is known in advance
CC and low probability of association with extraneous, possibly toxic,
CC genes. AA24894-244980 represent the Pseudomonas alcaligenes repeat (PAR)
CC elements described in the method of the invention.

SQ Sequence 77 BP; 19 A; 26 C; 17 G; 15 T; 0 other;

Query Match 96.1%; Score 74; DB 21; Length 77;
Best Local Similarity 100.0%; Pred. No. 1.1e-18;
Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 ctacacaaatggttcaagcgttcgctcactcgcggaacggcgttaagcgccctt 62
|||||
Db 3 ctacacaaatggttcaagcgttcgctcactcgcggaacggcgttaagcgccctt 62
|||||

QY 63 aaccaaagcgttaga 76
|||||
Db 63 aaccaaagcgttaga 76
|||||

RESULT 4
AAZ44981
ID AAZ44981 standard; DNA; 14143 BP.
XX
AC AAZ44981;
XX
DT 16-MAY-2000 (first entry)
XX
DE P. alcaligenes restriction enzyme PacI superintegron DNA #1.
XX
KW Diversity-selected gene; restriction enzyme; adhesin; toxin;
KW detoxifying enzyme; PacI; superintegron; ss.
XX
OS Pseudomonas alcaligenes.
XX
PN WO9964632-Al.
XX
PD 16-DEC-1999.
XX
PF 11-JUN-1999; 99WO-US13295.
XX
PR 12-JUN-1998; 98US-0089086.
XX
PR 12-JUN-1998; 98US-0089101.
XX
PA (NEWWE) NEW ENGLAND BIOLABS INC.
XX
PI Raleigh EA, Vaisvila R, Morgan RD;
XX
DR WPI; 2000-116558/10.
XX
PT Cloning intact genes used to isolate genes for restriction enzymes -
XX
PS Example 1A; Fig 3A-1; 97pp; English.

CC This invention describes a novel method for cloning intact,
CC diversity-selected genes (I) from within gene cassettes (GC) which
CC comprises identifying DNA repeats that flank GC, hybridizing
CC oligonucleotides (ON) to these repeats and amplification to produce DNA
CC fragments containing (I), ligating these fragments into a vector and
CC transforming cells with the vector. This method is used to clone a wide
CC variety of prokaryotic genes that provide a selective advantage under
CC particular conditions, particularly those that encode restriction
CC enzymes (used as reagents in molecular biology); adhesins (for use in
CC coating or for targeting molecules or organisms to particular sites, e.g.
CC for competitive exclusion of a selected pathogen); detoxifying enzymes;
CC toxins that interact with a host, e.g. for synthesis of inhibitors or

CC antagonists of the toxin, or in vaccination, or a modification
CC methyltransferase. Intact genes can be cloned directly with a high
CC probability that the orientation of expression is known in advance and
CC low probability of association with extraneous, possibly toxic, genes.
CC This sequence represents a Pseudomonas alcaligenes PacI superintegron
CC DNA sequence.

SQ Sequence 14143 BP; 3936 A; 3446 C; 3311 G; 3449 T; 1 other;

Query Match 96.1%; Score 74; DB 21; Length 14143;
Best Local Similarity 100.0%; Pred. No. 3.4e-18;
Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 ctacacaaatggttcaagcgttcgctcactcgcggaacggcgttaagcgccctt 62
|||||
Db 9095 ctacacaaatggttcaagcgttcgctcactcgcggaacggcgttaagcgccctt 9154
|||||

QY 63 aaccaaagcgttaga 76
|||||
Db 9155 aaccaaagcgttaga 9168
|||||

RESULT 5
AAZ44908
ID AAZ44908 standard; DNA; 77 BP.
XX
AC AAZ44908;
XX
DT 16-MAY-2000 (first entry)
XX
DE P. alcaligenes repeat (PAR) element DNA #15.
XX
KW Diversity-selected gene; restriction enzyme; adhesin; toxin;
KW detoxifying enzyme; repeat element; PAR; ss.
XX
OS Pseudomonas alcaligenes.
XX
PN WO9964632-Al.
XX
PD 16-DEC-1999.
XX
PF 11-JUN-1999; 99WO-US13295.
XX
PR 12-JUN-1998; 98US-0089086.
XX
PR 12-JUN-1998; 98US-0089101.
XX
PA (NEWWE) NEW ENGLAND BIOLABS INC.
XX
PI Raleigh EA, Vaisvila R, Morgan RD;
XX
DR WPI; 2000-116558/10.
XX
PT Cloning intact genes used to isolate genes for restriction enzymes -
XX
PS Claim 7a; Page 59; 97pp; English.

CC This invention describes a novel method for cloning intact,
CC diversity-selected genes (I) from within gene cassettes (GC) which
CC comprises identifying DNA repeats that flank GC, hybridizing
CC oligonucleotides (ON) to these repeats and amplification to produce DNA
CC fragments containing (I), ligating these fragments into a vector and
CC transforming cells with the vector. This method is used to clone a wide
CC variety of prokaryotic genes that provide a selective advantage under
CC particular conditions, particularly those that encode restriction
CC enzymes (used as reagents in molecular biology); adhesins (for use in
CC coating or for targeting molecules or organisms to particular sites,
CC e.g. for competitive exclusion of a selected pathogen); detoxifying
CC enzymes; toxins that interact with a host, e.g. for synthesis of
CC inhibitors or antagonists of the toxin, or in vaccination, or a
CC modification methyltransferase. Intact genes can be cloned directly with
CC a high probability that the orientation of expression is known in advance
CC and low probability of association with extraneous, possibly toxic,

CC genes. AAZ44894-244980 represent the Pseudomonas alcaligenes repeat (PAR)
 CC elements described in the method of the invention.

SQ Sequence 77 BP; 18 A; 26 C; 17 G; 16 T; 0 other;

Query Match 93.2%; Score 71.8; DB 21; Length 77;
 Best Local Similarity 97.3%; Pred. NO. 7.3e-18;
 Matches 73; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 ctctaaacaaatgttcaagcgttcgtctcgtcactcggcgacggtctaaagcggccccc 60
 |||||
 Db 1 ctctaaacaaatgttcaagcgttcgtctcgtcactcggcgacggtctaaagcggccccc 60
 |||||
 QY 61 ttaacccaaacgttag 75
 |||||
 Db 61 ttaacccaaacgttag 75

RESULT 6
 AAZ88511
 ID AAZ88511 standard; DNA; 74 BP.
 AC AAZ88511;
 XX
 DT 16-MAY-2000 (first entry)
 DE P. alcaligenes repeat (PAR) element DNA PARF8.
 XX
 KW Diversity-selected gene; restriction enzyme; adhesin; toxin;
 KW detoxifying enzyme; repeat element; PAR; ss.
 XX
 OS Pseudomonas alcaligenes.
 XX
 PN WO9964632-A1.
 XX
 PD 16-DEC-1999.
 XX
 PF 11-JUN-1999; 99WO-US13295.
 XX
 PR 12-JUN-1998; 98US-0089086.
 PR 12-JUN-1998; 98US-0089101.
 XX
 PA (NEWE) NEW ENGLAND BIOLABS INC.
 XX
 PI Raleigh EA, Vaisvila R, Morgan RD;
 XX
 DR WPI; 2000-116558/10.
 XX

Cloning intact genes used to isolate genes for restriction enzymes -

Example 1B; Fig 3E; 97pp; English.
 CC This invention describes a novel method for cloning intact,
 CC diversity-selected genes (I) from within gene cassettes (GC) which
 CC comprises identifying DNA repeats that flank GC, hybridizing
 CC oligonucleotides (ON) to these repeats and amplification to produce DNA
 CC fragments containing (I), ligating these fragments into a vector and
 CC transforming cells with the vector. This method is used to clone a wide
 CC variety of prokaryotic genes that provide a selective advantage under
 CC particular conditions, particularly those that encode restriction
 CC enzymes (used as reagents in molecular biology); adhesins (for use in
 CC coating or for targeting molecules or organisms to particular sites,
 CC e.g. for competitive exclusion of a selected pathogen); detoxifying
 CC enzymes; toxins that interact with a host, e.g. for synthesis of
 CC inhibitors or antagonists of the toxin, or in vaccination, or a
 CC modification methyltransferase. Intact genes can be cloned directly with
 CC and low probability of association with extraneous, possibly toxic,
 CC genes. AAZ88504-288521 represent Pseudomonas alcaligenes repeat (PAR)
 CC elements described in the method of the invention.

SQ Sequence 74 BP; 18 A; 23 C; 17 G; 16 T; 0 other;

Query Match 92.7%; Score 71.4; DB 21; Length 74;
 Best Local Similarity 98.6%; Pred. NO. 1e-17;
 Matches 72; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 ctacaaatgttcaagcgttcgtctcgtcactcggcgacggtctaaagcggccccc 62
 |||||
 Db 1 ctacaaatgttcaagcgttcgtctcgtcactcggcgacggtctaaagcggccccc 60
 |||||
 QY 63 aacccaaacgttag 75
 |||||
 Db 61 aacccaaacgttag 73

RESULT 7
 AAZ44952
 ID AAZ44952 standard; DNA; 77 BP.
 AC AAZ44952;
 XX
 DT 16-MAY-2000 (first entry)
 DE P. alcaligenes repeat (PAR) element DNA #59.
 XX
 KW Diversity-selected gene; restriction enzyme; adhesin; toxin;
 KW detoxifying enzyme; repeat element; PAR; ss.
 XX
 OS Pseudomonas alcaligenes.
 XX
 PN WO9964632-A1.
 XX
 PD 16-DEC-1999.
 XX
 PF 11-JUN-1999; 99WO-US13295.
 XX
 PR 12-JUN-1998; 98US-0089086.
 PR 12-JUN-1998; 98US-0089101.
 XX
 PA (NEWE) NEW ENGLAND BIOLABS INC.
 XX
 PI Raleigh EA, Vaisvila R, Morgan RD;
 XX
 DR WPI; 2000-116558/10.
 XX

Cloning intact genes used to isolate genes for restriction enzymes -
 Claim 7a; Page 60; 97pp; English.

CC This invention describes a novel method for cloning intact,
 CC diversity-selected genes (I) from within gene cassettes (GC) which
 CC comprises identifying DNA repeats that flank GC, hybridizing
 CC oligonucleotides (ON) to these repeats and amplification to produce DNA
 CC fragments containing (I), ligating these fragments into a vector and
 CC transforming cells with the vector. This method is used to clone a wide
 CC variety of prokaryotic genes that provide a selective advantage under
 CC particular conditions, particularly those that encode restriction
 CC enzymes (used as reagents in molecular biology); adhesins (for use in
 CC coating or for targeting molecules or organisms to particular sites,
 CC e.g. for competitive exclusion of a selected pathogen); detoxifying
 CC enzymes; toxins that interact with a host, e.g. for synthesis of
 CC inhibitors or antagonists of the toxin, or in vaccination, or a
 CC modification methyltransferase. Intact genes can be cloned directly with
 CC and high probability of association with extraneous, possibly toxic,
 CC genes. AAZ44894-244980 represent the Pseudomonas alcaligenes repeat (PAR)
 CC elements described in the method of the invention.

SQ Sequence 77 BP; 18 A; 25 C; 18 G; 16 T; 0 other;

Query Match 92.7%; Score 71.4; DB 21; Length 77;
 Best Local Similarity 98.6%; Pred. NO. 1e-17;

Matches 72; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 ctacaaatggttcaagcgttcgtctcactcgtggaacgcggtctaaagccgcctt 62
 |||||
 Db 3 ctacaaatggttcaagcgttcgtctcactcgtggaacgcggtctaaagccgcctt 62
 |||||
 QY 63 aaccaaacttag 75
 |||||
 Db 63 aaccaaacttag 75

RESULT 8
 ID AAZ44987 standard; DNA; 77 BP.
 XX
 AC AAZ44987;
 XX
 DT 16-MAY-2000 (first entry)
 XX
 DE P. alcaligenes repeat (PAR) element family 1 consensus DNA #3.
 XX
 Diversity-selected gene; restriction enzyme; adhesin; toxin;
 KW detoxifying enzyme; repeat element; PAR; family 1; ss.
 XX
 OS Pseudomonas alcaligenes.
 XX
 PN WO9964632-A1.
 XX
 PD 16-DEC-1999.
 XX
 PF 11-JUN-1999; 99WO-US13295.
 XX
 PR 12-JUN-1998; 98US-0089086.
 XX
 PR 12-JUN-1998; 98US-0089101.
 XX
 PA (NEW) NEW ENGLAND BIOLABS INC.
 XX
 PI Raleigh EA, Vaisvila R, Morgan RD;
 XX
 DR WPI; 2000-116558/10.
 XX
 PT Cloning intact genes used to isolate genes for restriction enzymes -
 XX
 PS Example 1B; Fig 6A; 97pp; English.

This invention describes a novel method for cloning intact, diversity-selected genes (I) from within gene cassettes (GC) which comprises identifying DNA repeats that flank GC, hybridizing oligonucleotides (ON) to these repeats and amplification to produce DNA fragments containing (I), ligating these fragments into a vector and transforming cells with the vector. This method is used to clone a wide variety of prokaryotic genes that provide a selective advantage under particular conditions, particularly those that encode restriction enzymes (used as reagents in molecular biology); adhesins (for use in coating or for targeting molecules or organisms to particular sites, e.g. for competitive exclusion of a selected pathogen); detoxifying enzymes; toxins that interact with a host, e.g. for synthesis of inhibitors or antagonists of the toxin, or in vaccination, or a modification methyltransferase. Intact genes can be cloned directly with a high probability that the orientation of expression is known in advance and low probability of association with extraneous, possibly toxic, genes. AAZ44987 represents Pseudomonas alcaligenes repeat (PAR) element family 1 consensus sequences described in the method of the invention.

Sequence 77 BP; 17 A; 26 C; 18 G; 16 T; 0 other;

Query Match 92.7%; Score 71.4; DB 21; Length 77;
 Best Local Similarity 98.6%; Pred. No. 1e-17;
 Matches 72; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 ctacaaatggttcaagcgttcgtctcactcgtggaacgcggtctaaagccgcctt 62

Db 3 ctacaaatggttcaagcgttcgtctcactcgtggaacgcggtctaaagccgcctt 62
 |||||
 QY 63 aaccaaacttag 75
 |||||
 Db 63 aaccaaacttag 75

RESULT 9
 ID AAZ44954 standard; DNA; 77 BP.
 XX
 AC AAZ44954;
 XX
 DT 16-MAY-2000 (first entry)
 XX
 DE P. alcaligenes repeat (PAR) element DNA #61.
 XX
 Diversity-selected gene; restriction enzyme; adhesin; toxin;
 KW detoxifying enzyme; repeat element; PAR; ss.
 XX
 OS Pseudomonas alcaligenes.
 XX
 PN WO9964632-A1.
 XX
 PD 16-DEC-1999.
 XX
 PF 11-JUN-1999; 99WO-US13295.
 XX
 PR 12-JUN-1998; 98US-0089086.
 XX
 PR 12-JUN-1998; 98US-0089101.
 XX
 PA (NEW) NEW ENGLAND BIOLABS INC.
 XX
 PI Raleigh EA, Vaisvila R, Morgan RD;
 XX
 DR WPI; 2000-116558/10.
 XX
 PT Cloning intact genes used to isolate genes for restriction enzymes -
 XX
 PS Claim 7a; Page 60; 97pp; English.

This invention describes a novel method for cloning intact, diversity-selected genes (I) from within gene cassettes (GC) which comprises identifying DNA repeats that flank GC, hybridizing oligonucleotides (ON) to these repeats and amplification to produce DNA fragments containing (I), ligating these fragments into a vector and transforming cells with the vector. This method is used to clone a wide variety of prokaryotic genes that provide a selective advantage under particular conditions, particularly those that encode restriction enzymes (used as reagents in molecular biology); adhesins (for use in coating or for targeting molecules or organisms to particular sites, e.g. for competitive exclusion of a selected pathogen); detoxifying enzymes; toxins that interact with a host, e.g. for synthesis of inhibitors or antagonists of the toxin, or in vaccination, or a modification methyltransferase. Intact genes can be cloned directly with a high probability that the orientation of expression is known in advance and low probability of association with extraneous, possibly toxic, genes. AAZ44954 represents the Pseudomonas alcaligenes repeat (PAR) elements described in the method of the invention.

Sequence 77 BP; 18 A; 25 C; 18 G; 16 T; 0 other;

Query Match 92.5%; Score 71.2; DB 21; Length 77;
 Best Local Similarity 96.1%; Pred. No. 1.2e-17;
 Matches 73; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 tctacaaatggttcaagcgttcgtctcactcgtggaacgcggtctaaagccgcctt 61
 |||||
 Db 2 tctacaaatggttcaagcgttcgtctcactcgtggaacgcggtctaaagccgcctt 61
 |||||
 QY 62 taaccaaacttag 77

RESULT	11
AAZ44913	
ID	AAZ44913 standard; DNA; 77 BP.
XX AC	
AAZ44913;	
XX DT	16-MAY-2000 (first entry)
XX DE	P. alcaligenes repeat (PAR) element DNA #20.
XX KW	Diversity-selected gene; restriction enzyme; adhesin; toxin;
KW decoupling enzyme; repeat element; PAR; ss.	
XX OS	Pseudomonas alcaligenes.
OS PN	WO9964632-A1.
PN XX	16-DEC-1999.
PD XX	11-JUN-1999; 99WO-US13295.
PF XX	12-JUN-1998; 98US-0089086.
PR XX	12-JUN-1998; 98US-0089101.
PA XX	(NEWWE) NEW ENGLAND BIOLABS INC.
PI XX	Raleigh EA, Vaisvila R, Morgan RD;
DR XX	WPI; 2000-116558/10.
PT XX	Cloning intact genes used to isolate genes for restriction enzymes -
PS Claim 7a; Page 59; 97pp; English.	
CC This invention describes a novel method for cloning intact,	
CC diversity-selected genes (I) from within gene cassettes (GC) which	
CC comprises identifying DNA repeats that flank GC, hybridizing	
CC oligonucleotides (ON) to these repeats and amplification to produce DNA	
CC fragments containing (I), ligating these fragments into a vector and	
CC transforming cells with the vector. This method is used to clone a wide	
CC variety of prokaryotic genes that provide a selective advantage under	
CC particular conditions, particularly those that encode restriction	
CC enzymes (used as reagents in molecular biology); adhering (for use in	
CC coating or for targeting molecules or organisms to particular sites,	
CC e.g. for competitive exclusion of a selected pathogen); detoxifying	
CC enzymes; toxins that interact with a host, e.g. for synthesis of	
CC inhibitors or antagonists of the toxin, or in vaccination, or a	
CC modification methyltransferase. Intact genes can be cloned directly with	
CC a high probability that the orientation of expression is known in advance	
CC and low probability of association with extraneous, possibly toxic,	
CC genes. AAZ44894-744980 represent the Pseudomonas alcaligenes repeat (PAR)	
CC elements described in the method of the invention.	
SQ Sequence 77 BP; 18 A; 25 C; 17 G; 17 T; 0 other;	
Query Match	90.6%; Score 69.8; DB 21; Length 77;
Best Local Similarity	97.3%; Pred. No. 4.1e-17;
Matches 71; Conservative	0; Mismatches 2; Indels 0; Gaps 0
Qy 3 ctacaataatggttcaagcggcttcggtccacctacgcggaccggttaaacgccgccttt 62	
Db	
3 ctacaataatggttcaagcggcttcggtccacctacgcggaccggttaaacgccgccttt 62	
Qy 63 aaccaaacgtag 75	
Db	
63 aaccaaacgtag 75	
RESULT 12	
AAZ44944	
ID ID	AAZ44944 standard; DNA; 77 BP.
XX	


```

XX OS Pseudomonas alcaligenes.
XX PN WO9964632-A1.
XX PD 16-DEC-1999.
XX PF 11-JUN-1999; 99WO-US13295.
XX PR 12-JUN-1998; 98US-0089086.
XX PA (NEW ) NEW ENGLAND BIOLABS INC.
XX PI Raleigh EA, Vaisvila R, Morgan RD;
XX WPI; 2000-116558/10.
XX PT Cloning intact genes used to isolate genes for restriction enzymes
XX PS Example 1B; Fig 3E; 97pp; English.
XX CC This invention describes a novel method for cloning intact,
XX CC diversity-selected genes (I) from within gene cassettes (GC) which
XX CC comprises identifying DNA repeats that flank GC, hybridizing
XX CC oligonucleotides (ON) to these repeats and amplification to produce DNA
XX CC fragments containing (I), ligating these fragments into a vector and
XX CC transforming cells with the vector. This method is used to clone a wide
XX CC variety of prokaryotic genes that provide a selective advantage under
XX CC particular conditions, particularly those that encode restriction
XX CC enzymes (used as reagents in molecular biology); adhesins (for use in
XX CC coating or for targeting molecules or organisms to particular sites,
XX CC e.g. for competitive exclusion of a selected pathogen); detoxifying
XX CC inhibitors or antagonists of the toxin, or in vaccination, or a
XX CC modification methyltransferase. Intact genes can be cloned directly with
XX CC a high probability that the orientation of expression is known in advance
XX CC and low probability of association with extraneous, possibly toxic,
XX CC genes. AAZ88504-288521 represent Pseudomonas alcaligenes repeat (PAR)
XX CC elements described in the method of the invention.
XX SQ Sequence 74 BP; 16 A; 25 C; 16 G; 17 T; 0 other;

Query Match 88.6%; Score 68.2; DB 21; Length 74;
Best Local Similarity 95.9%; Pred. No. 1.6e-16;
Matches 70; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 ctaacaaatggttcaagcgttcgtcgtcactcgtggaccggtctaaagccggccctt 62
Db 1 ctaactactggttcaagtcgttcgtcgtcactcgtggaccggtctaaagccggccctt 60

QY 63 aaccaaacgttag 75
Db 61 aaccaaacgttag 73

RESULT 15
AAZ88515
ID AAZ88515 standard; DNA; 74 BP.
XX AC AAZ88515;
XX DT 16-MAY-2000 (first entry)
XX DE P. alcaligenes repeat (PAR) element DNA PARf12.
XX KW Diversity-selected gene; restriction enzyme; adhesin; toxin;
XX KW detoxifying enzyme; repeat element; PAR; ss.
XX OS Pseudomonas alcaligenes.
XX PN WO9964632-A1.

```

```

XX PD 16-DEC-1999.
XX PF 11-JUN-1999; 99WO-US13295.
XX PR 12-JUN-1998; 98US-0089086.
XX PA (NEW ) NEW ENGLAND BIOLABS INC.
XX PI Raleigh EA, Vaisvila R, Morgan RD;
XX WPI; 2000-116558/10.
XX PT Cloning intact genes used to isolate genes for restriction enzymes
XX PS Example 1B; Fig 3E; 97pp; English.
XX CC This invention describes a novel method for cloning intact,
XX CC diversity-selected genes (I) from within gene cassettes (GC) which
XX CC comprises identifying DNA repeats that flank GC, hybridizing
XX CC oligonucleotides (ON) to these repeats and amplification to produce DNA
XX CC fragments containing (I), ligating these fragments into a vector and
XX CC transforming cells with the vector. This method is used to clone a wide
XX CC variety of prokaryotic genes that provide a selective advantage under
XX CC particular conditions, particularly those that encode restriction
XX CC enzymes (used as reagents in molecular biology); adhesins (for use in
XX CC coating or for targeting molecules or organisms to particular sites,
XX CC e.g. for competitive exclusion of a selected pathogen); detoxifying
XX CC inhibitors or antagonists of the toxin, or in vaccination, or a
XX CC modification methyltransferase. Intact genes can be cloned directly with
XX CC a high probability that the orientation of expression is known in advance
XX CC and low probability of association with extraneous, possibly toxic,
XX CC genes. AAZ88504-288521 represent Pseudomonas alcaligenes repeat (PAR)
XX CC elements described in the method of the invention.
XX SQ Sequence 74 BP; 16 A; 24 C; 17 G; 17 T; 0 other;

Query Match 88.6%; Score 68.2; DB 21; Length 74;
Best Local Similarity 95.9%; Pred. No. 1.6e-16;
Matches 70; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 ctaacaaatggttcaagcgttcgtcgtcactcgtggaccggtctaaagccggccctt 62
Db 1 ctaactactggttcaagtcgttcgtcgtcactcgtggaccggtctaaagccggccctt 60

QY 63 aaccaaacgttag 75
Db 61 aaccaaacgttag 73

Search completed: August 2, 2001, 16:57:28
Job time: 258 sec

```



```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:      August 2, 2001, 16:55:10 ; Search time 89.09 Seconds
              (without alignments)
              163.621 Million cell updates/sec

Title:       US-09-701-626A-78
Perfect score: 77
Sequence:    1 ctctacaatatggtcaagc.....cccttaaccaacqtagag 77

Scoring table:
IDENTITY_NUC
Gapox 10.0 , Gapext 1.0

Searched:   324599 seqs, 94655562 residues

```


TELECOMMUNICATION INFORMATION:
TELEPHONE: (908) 594-3903
TELEFAX: (908) 594-4720
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3870 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-138-133-1

Query Match 29.1%; Score 22.4; DB 1; Length 3870;
Best Local Similarity 62.5%; Pred. No. 16;
Matches 35; Conservative 0; Mismatches 21; Indels 0; Gaps 0;
Db 5 acaaatggttcaagcgttcgctcactcgggacggcgctaaagcggcccc 60
||||| ||| ||||| || ||| ||||| || ||| || ||||| |
391 AACATGTGGATCAAGGCTTAACTTGCTCATGGAAGATACGCTGCAGCGGCCAC 446

RESULT 5
US-08-138-641-3
Sequence 3, Application US/08138641
Patent No. 5474921
GENERAL INFORMATION:
APPLICANT: Koblán, Kenneth S.
APPLICANT: Pompliano, David L.
TITLE OF INVENTION: ASSAY TO DETERMINE INHIBITORS OF
OPERATING SYSTEM: PC-DOS/MS-DOS
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: David A. Muthard
STREET: P.O. Box 2000, 126 E. Lincoln Avenue
CITY: Rahway
STATE: New Jersey
COUNTRY: USA
ZIP: 07065
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/138,641
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Muthard, David A.
REGISTRATION NUMBER: 35,297
REFERENCE/DOCKET NUMBER: 18937
TELECOMMUNICATION INFORMATION:
TELEPHONE: (908) 594-3903
TELEFAX: (908) 594-4720
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 3893 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-138-641-3

Query Match 29.1%; Score 22.4; DB 1; Length 3893;
Best Local Similarity 62.5%; Pred. No. 16;
Matches 35; Conservative 0; Mismatches 21; Indels 0; Gaps 0;
Db 5 acaaatggttcaagcgttcgctcactcgggacggcgctaaagcggcccc 60
||||| ||| ||||| || ||| ||||| || ||| || ||||| |
391 AACATGTGGATCAAGGCTTAACTTGCTCATGGAAGATACGCTGCAGCGGCCAC 446

RESULT 6
US-08-138-133-3
Sequence 3, Application US/08138133
Patent No. 5519163
GENERAL INFORMATION:
APPLICANT: GIBBS, JACKSON B.
APPLICANT: KOBLAN, KENNETH S.
APPLICANT: MACLEOD, ANGUS M.
APPLICANT: MERCHANT, KEVIN J.
TITLE OF INVENTION: INHIBITORS OF PHOSPHOINOSITIDE-SPECIFIC
CORRESPONDENCE ADDRESS:
NUMBER OF SEQUENCES: 9
ADDRESSEE: DAVID A. MUTHARD
STREET: P.O. BOX 2000, 126 E. LINCOLN AVENUE
CITY: RAHWAY
STATE: NEW JERSEY
COUNTRY: U.S.A.
ZIP: 07065
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/138,133
FILING DATE:
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: MUTHARD, DAVID A.
REGISTRATION NUMBER: 35,297
REFERENCE/DOCKET NUMBER: 18938
TELECOMMUNICATION INFORMATION:
TELEPHONE: (908) 594-3903
TELEFAX: (908) 594-4720
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 3893 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-138-133-3

Query Match 29.1%; Score 22.4; DB 1; Length 3893;
Best Local Similarity 62.5%; Pred. No. 16;
Matches 35; Conservative 0; Mismatches 21; Indels 0; Gaps 0;
QY 5 acaaatggttcaagcgttcgctcactcgggacggcgctaaagcggcccc 60
||||| ||| ||||| || ||| ||||| || ||| || ||||| |
Db 391 AACATGTGGATCAAGGCTTAACTTGCTCATGGAAGATACGCTGCAGCGGCCAC 446

RESULT 7
US-08-738-172-3
Sequence 3, Application US/08738172
Patent No. 5939257
GENERAL INFORMATION:
APPLICANT: Szasz, Joseph
APPLICANT: Davis, Maria
TITLE OF INVENTION: THERMOSTABLE ALKALINE PHOSPHATASES
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lyon & Lyon
STREET: 633 West Fifth Street
CITY: Los Angeles
STATE: California
COUNTRY: U.S.A.
ZIP: 90071-2066
COMPUTER READABLE FORM:

;; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
;; MEDIUM TYPE: storage
;; COMPUTER: IBM Compatible
;; OPERATING SYSTEM: IBM P.C. DOS 5.0
;; SOFTWARE: Word Perfect 5.1
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/738,172
;; FILING DATE:
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 60/005,965
;; FILING DATE: October 27, 1995
;; APPLICATION NUMBER: 08/465,003
;; FILING DATE: June 5, 1995
;; APPLICATION NUMBER: 08/575,354
;; FILING DATE: December 20, 1995
;; APPLICATION NUMBER: 08/240,158
;; FILING DATE: May 10, 1994
;; APPLICATION NUMBER: 08/229,329
;; FILING DATE: April 18, 1994
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Warburg, Richard J.
;; REGISTRATION NUMBER: 32,327
;; REFERENCE/DOCKET NUMBER: 222/158
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (213) 489-1600
;; TELEFAX: (213) 955-0440
;; TELEX: 67-3510
;; INFORMATION FOR SEQ ID NO: 3:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1368 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; US-08-738-172-3

Query Match 28.8%; Score 22.2; DB 2; Length 1368;
Best Local Similarity 61.0%; Pred. No. 15;
Matches 36; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 2 tctacaagaatgttcaagcgttcctcactcgggacgcggtataagcggcccc 60
||| | |||| | | | ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 752 TCTTCTGATGGTGGAGGCGCGCATCGACACGCGCGGACATGCCAACGCGCGCC 810

RESULT 8
US-08-738-172-1/c
; Sequence 1, Application US/08738172
; Patent No. 5939257
; GENERAL INFORMATION:
; APPLICANT: Szasz, Joseph
; APPLICANT: Davis, Maria
; TITLE OF INVENTION: THERMOSTABLE ALKALINE PHOSPHATASES
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; STREET: Suite 4700
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071-2066
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/738,172
; FILING DATE:
; CLASSIFICATION: 435

;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 60/005,965
;; FILING DATE: October 27, 1995
;; APPLICATION NUMBER: 08/465,003
;; FILING DATE: June 5, 1995
;; APPLICATION NUMBER: 08/575,354
;; FILING DATE: December 20, 1995
;; APPLICATION NUMBER: 08/240,158
;; FILING DATE: May 10, 1994
;; APPLICATION NUMBER: 08/229,329
;; FILING DATE: April 18, 1994
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Warburg, Richard J.
;; REGISTRATION NUMBER: 32,327
;; REFERENCE/DOCKET NUMBER: 222/158
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (213) 489-1600
;; TELEFAX: (213) 955-0440
;; TELEX: 67-3510
;; INFORMATION FOR SEQ ID NO: 1:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 2936 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; US-08-738-172-1

Query Match 28.8%; Score 22.2; DB 2; Length 2936;
Best Local Similarity 61.0%; Pred. No. 18;
Matches 36; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 2 tctacaagaatgttcaagcgttcctcactcgggacgcggtataagcggcccc 60
||| | |||| | | | ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 791 TCTTCTGATGGTGGAGGCGCGCATCGACACGCGCGGACATGCCAACGCGCGCC 733

RESULT 9
US-07-952-853-21/c
; Sequence 21, Application US/07952853
; Patent No. 5863783
; GENERAL INFORMATION:
; APPLICANT: Van Heuvel, Margaretha
; APPLICANT: Bakhuis, Janna G.
; APPLICANT: Coutel, Yves
; APPLICANT: Harder, Abraham
; APPLICANT: De Graaff, Leendert H.
; APPLICANT: Flippin, Michel J. A.
; APPLICANT: Van Der Veen, Peter
; APPLICANT: Visser, Jacob
; APPLICANT: Andreoli, Peter M.
; TITLE OF INVENTION: CLONING AND EXPRESSION OF DNA
; TITLE OF INVENTION: MOLECULES
; TITLE OF INVENTION: ENCODING ARABINAN-DEGRADING ENZYMES OF FUNGAL
; TITLE OF INVENTION: ORIGIN
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 755 Page Mill Road
; CITY: Palo Alto
; STATE: California
; ZIP: 94304-1018
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/952,853
; FILING DATE: 19921125
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Murashige, Kate H.

```
;
;
; REGISTRATION NUMBER: 29,959
; REFERENCE/DOCKET NUMBER: 246152003500
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-813-5600
; TELEFAX: 415-494-0792
; TELEX: 706141
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3958 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Aspergillus niger
; FEATURE:
; NAME/KEY: exon
; LOCATION: 1247..1390
; FEATURE:
; NAME/KEY: intron
; LOCATION: 1391..1442
; FEATURE:
; NAME/KEY: exon
; LOCATION: 1443..1957
; FEATURE:
; NAME/KEY: intron
; LOCATION: 1958..2005
; FEATURE:
; NAME/KEY: exon
; LOCATION: 2006..2089
; FEATURE:
; NAME/KEY: intron
; LOCATION: 2090..2137
; FEATURE:
; NAME/KEY: exon
; LOCATION: 2138..2214
; FEATURE:
; NAME/KEY: intron
; LOCATION: 2215..2262
; FEATURE:
; NAME/KEY: exon
; LOCATION: 2263..2295
; FEATURE:
; NAME/KEY: intron
; LOCATION: 2296..2346
; FEATURE:
; NAME/KEY: exon
; LOCATION: 2347..2498
; FEATURE:
; NAME/KEY: intron
; LOCATION: 2499..2548
; FEATURE:
; NAME/KEY: exon
; LOCATION: 2549..3037
; FEATURE:
; NAME/KEY: intron
; LOCATION: 3038..3092
; FEATURE:
; NAME/KEY: exon
; LOCATION: 3093..3485
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Join(1247..1390, 1443..1957, 2006..2089,
; LOCATION: 2138..2214, 2263..2295, 2347..2498, 2549..3037,
; LOCATION: 3093..3485)
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /codon_start= 1247
; OTHER INFORMATION: /product= "alpha-L-arabinofuranosidase A"
; OTHER INFORMATION: /evidence= EXPERIMENTAL
; OTHER INFORMATION: /gene= "exoA"
; FEATURE:
```

```
;
; NAME/KEY: sig_peptide
; LOCATION: 1247..1321
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 1322..3485
; US-07-952-853-21

Query Match 28.8%; Score 22.2; DB 2; Length 3958;
Best Local Similarity 58.2%; Pred. No. 20;
Matches 39; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

QY 11 tggttcaagccgttcgcttcgctcactcggagaccgctaaagccgccccttaaccacaac 70
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 304 TGGTTTACCCGGCTTAATCTCTCCGGGAAGTCGTGCCACCTTCGTATATGCCCAACT 245

QY 71 gttagag 77
||||| |

Db 244 GTTAGAG 238

RESULT 10
US-08-914-848-21/c
; Sequence 21, Application US/08914848
; Patent No. 5989887
; GENERAL INFORMATION:
; APPLICANT: Van Heuvel, Margaretha
; APPLICANT: Bakhuis, Janna G.
; APPLICANT: Coutel, Yves
; APPLICANT: Harder, Abraham
; APPLICANT: De Graaff, Leendert H.
; APPLICANT: Flippi, Michel J. A.
; APPLICANT: Van Der Veen, Peter
; APPLICANT: Visser, Jacob
; APPLICANT: Andreoli, Peter M.
; TITLE OF INVENTION: CLONING AND EXPRESSION OF DNA
; TITLE OF INVENTION: MOLECULES ENCODING ARABINAN-DEGRADING ENZYMES OF FUNGAL
; TITLE OF INVENTION: ORIGIN
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 755 Page Mill Road
; CITY: Palo Alto
; STATE: California
; ZIP: 94304-1018
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/914,848
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/952,853
; FILING DATE: 25-NOV-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Murashige, Kate H.
; REGISTRATION NUMBER: 29,959
; REFERENCE/DOCKET NUMBER: 246152003500
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-813-5600
; TELEFAX: 415-494-0792
; TELEX: 706141
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3958 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
```



```
; Sequence 1, Application US/08832535
; Patent No. 5919658
; GENERAL INFORMATION:
; APPLICANT: NI, JIAN
; APPLICANT: LI, HAODONG
; APPLICANT: YU, GUO-LIANG
; APPLICANT: GENTZ, REINER L
; TITLE OF INVENTION: HUMAN CYSTATIN F
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HUMAN GENOME SCIENCES, INC.
; STREET: 9410 KEY WEST AVENUE
; CITY: ROCKVILLE
; STATE: MD
; COUNTRY: US
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/832,535
; FILING DATE: 03-APR-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: KIMBALL, PAUL C.
; REGISTRATION NUMBER: 34,610
; REFERENCE/DOCKET NUMBER: PF265
; TELEPHONE: (201) 994-1700
; TELEFAX: (201) 994-1744
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 633 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 67..501
; US-08-832-535-1

Query Match      28.6%; Score 22; DB 2; Length 633;
Best Local Similarity 59.7%; Pred. No. 14;
Matches 37; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

QY 1 ctctaacaaatggttcaagccgttcgctcactcgagaccggtctaaagccgcccc 60
   ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
Db 339 CTCGACATATATTTCAGGCCCTTCACTATCTGAAC TAGGCCCTTGTGTATGCGGACTC 280

QY 61 tt 62
   |
Db 279 CT 278

RESULT 13
US-09-019-485-1/c
; Sequence 1, Application US/09019485
; Patent No. 6066617
; GENERAL INFORMATION:
; APPLICANT: Li, Haodong
; APPLICANT: Yu, Guo-Liang
; APPLICANT: GENTZ, Reiner
; APPLICANT: Ni, Jian
; TITLE OF INVENTION: Cystatin F
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
```

```
; STATE: MD
; COUNTRY: US
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/019,485
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Benson, Robert H.
; REGISTRATION NUMBER: 30,446
; REFERENCE/DOCKET NUMBER: PF265P1
; TELEPHONE: 3013098504
; TELEFAX: 3013098439
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 633 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 67..501
; NAME/KEY: sig_peptide
; LOCATION: 67..123
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 124..501
; US-09-019-485-1

Query Match      28.6%; Score 22; DB 3; Length 633;
Best Local Similarity 59.7%; Pred. No. 14;
Matches 37; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

QY 1 ctctaacaaatggttcaagccgttcgctcactcgagaccggtctaaagccgcccc 60
   ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
Db 339 CTCGACATATATTTCAGGCCCTTCACTATCTGAAC TAGGCCCTTGTGTATGCGGACTC 280

QY 61 tt 62
   |
Db 279 CT 278

RESULT 14
US-08-791-522-2/c
; Sequence 2, Application US/08791522
; Patent No. 5935817
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Goli, Surya K.
; TITLE OF INVENTION: NOVEL HUMAN CYSTATIN-LIKE
; TITLE OF INVENTION: PROTEIN
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
```


